A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP4438.

# Mortality Risk and Fine Particulate Air Pollution in a Large, Representative Cohort of U.S. Adults

*C. Arden Pope III,[1] Jacob S. Lefler,[1] Majid Ezzati,[2] Joshua D. Higbee,[1] Julian D. Marshall,[3] Sun-Young Kim,[4] Matthew Bechle,[3] Kurtis S. Gilliat,[5] Spencer E. Vernon,[6] Allen L. Robinson,[7] and Richard T. Burnett[8]*

[1]Department of Economics, Brigham Young University, Provo, Utah, USA
[2]MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK
[3]Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington, USA
[4]Department of Cancer Control and Population Health, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang-si, Gyeonggi-do, Korea
[5]Center for the Economics of Human Development, University of Chicago, Chicago, Illinois, USA
[6]Cornerstone Research, San Francisco, California, USA
[7]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[8]Health Canada, Ottawa, Ontario, Canada

**Background:** Evidence indicates that air pollution contributes to cardiopulmonary mortality. There is ongoing debate regarding the size and shape of the pollution–mortality exposure–response relationship. There are also growing appeals for estimates of pollution–mortality relationships that use public data and are based on large, representative study cohorts.

**Objectives:** Our goal was to evaluate fine particulate matter air pollution ($PM_{2.5}$) and mortality using a large cohort that is representative of the U.S. population and is based on public data. Additional objectives included exploring model sensitivity, evaluating relative effects across selected subgroups, and assessing the shape of the $PM_{2.5}$–mortality relationship.

**Methods:** National Health Interview Surveys (1986–2014), with mortality linkage through 2015, were used to create a cohort of 1,599,329 U.S. adults and a subcohort with information on smoking and body mass index (BMI) of 635,539 adults. Data were linked with modeled ambient $PM_{2.5}$ at the census-tract level. Cox proportional hazards models were used to estimate $PM_{2.5}$–mortality hazard ratios for all-cause and specific causes of death while controlling for individual risk factors and regional and urban versus rural differences. Sensitivity and subgroup analyses were conducted and the shape of the $PM_{2.5}$–mortality relationship was explored.

**Results:** Estimated mortality hazard ratios, per $10\,\mu g/m^3$ long-term exposure to $PM_{2.5}$, were 1.12 (95% CI: 1.08, 1.15) for all-cause mortality, 1.23 (95% CI: 1.17, 1.29) for cardiopulmonary mortality, and 1.12 (95% CI: 1.00, 1.26) for lung cancer mortality. In general, $PM_{2.5}$–mortality associations were consistently positive for all-cause and cardiopulmonary mortality across key modeling choices and across subgroups of sex, age, race-ethnicity, income, education levels, and geographic regions.

**Discussion:** This large, nationwide, representative cohort of U.S. adults provides robust evidence that long-term $PM_{2.5}$ exposure contributes to cardiopulmonary mortality risk. The ubiquitous and involuntary nature of exposures and the broadly observed effects across subpopulations underscore the public health importance of breathing clean air. https://doi.org/10.1289/EHP4438

## Introduction

Epidemiological and related evidence implicates exposure to fine particulate matter ($PM_{2.5}$, particles <2.5 μm in aerodynamic diameter) air pollution as contributing to cardiopulmonary disease (Brook et al. 2010), lung cancer (Hamra et al. 2014), and infant mortality (Woodruff et al. 2006). Recent reports indicate that air pollution is also an important contributor to the global burden of disease (GBD 2015 Risk Factors Collaborators 2016; Cohen et al. 2017). $PM_{2.5}$ is largely generated (directly and indirectly) by the combustion of coal, diesel, gasoline, biofuels, and related high-temperature processes. $PM_{2.5}$ contains highly complex mixtures of particles, including soot, organics, nitrates, sulfates, and related chemicals that can penetrate deeply into the lungs.

Cohort studies have evaluated mortality risk associated with long-term exposure to $PM_{2.5}$ air pollution in the United States (Dockery et al. 1993; Pope et al. 2002, 2015, 2018; Miller et al. 2007; Puett et al. 2011; Lipsett et al. 2011; Lepeule et al. 2012; Hart et al. 2015; Thurston et al. 2016; Jerrett et al. 2017; Di et al. 2017; Parker et al. 2018), Canada (Crouse et al. 2015; Villeneuve et al. 2015; Pinault et al. 2016, 2017), Europe (Carey et al. 2013; Cesaroni et al. 2013; Beelen et al. 2014; Fischer et al. 2015; Bentayeb et al. 2015), and Asia (Tseng et al. 2015; Yin et al. 2017). A recent meta-analysis (Vodonos et al. 2018) indicated that these studies provide compelling evidence that long-term exposure to $PM_{2.5}$ contributes to increased risk of mortality. Study and cohort differences, however, make it difficult to estimate the overall mortality impacts representative of the entire U.S. adult population. For example, the pioneering Six Cities Study (Dockery et al. 1993; Lepeule et al. 2012) was designed to evaluate pollution-related effects in a representative cohort of adults—but a small cohort representing only six cities. The larger and more geographically representative American Cancer Society, Cancer Prevention Study II (ACS CPS-II) cohort (Pope et al. 2002, 2015; Jerrett et al. 2017) observed statistically robust $PM_{2.5}$–mortality associations, but the cohort overrepresented affluent, white, well-educated adults and the associations were smaller. Studies of other cohorts have been restricted to postmenopausal women in urban areas (Miller et al. 2007), health professionals (Puett et al. 2011; Hart et al. 2015), school teachers (Lipsett et al. 2011), or the elderly (Thurston et al. 2016). The largest cohort used to study $PM_{2.5}$–mortality associations included nearly 70 million U.S. Medicare beneficiaries (Di et al. 2017) and observed statistically robust $PM_{2.5}$–mortality associations. The Medicare cohort, however, included only elderly adults, was unable to analyze specific causes of death, and had limited ability to directly control for smoking status.

Address correspondence to C. Arden Pope III, 142 FOB, Brigham Young University, Provo, UT 84602 USA. Telephone: (801) 422-2157. Email: cap3@byu.edu

Supplemental Material is available online (https://doi.org/10.1289/EHP4438).

The authors declare they have no actual or potential competing financial interests.

Received 6 September 2018; Revised 9 May 2019; Accepted 22 May 2019; Published 24 July 2019; Corrected 27 September 2019.

**Note to readers with disabilities:** *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

The primary objective of the present analysis was to evaluate cause-specific $PM_{2.5}$–mortality associations in a large, contemporary, nationally representative cohort of U.S. adults based on open-use public data. Additional objectives included evaluating modeling sensitivity; exploring relative effects across age, sex, race-ethnicity, smoking status, education levels, and other subgroups; and flexibly assessing the shape of the $PM_{2.5}$–mortality relationship.

## Methods

### *Study Subjects and Data Access*

Study subjects comprised individuals 18–84 y of age living in the contiguous United States who were interviewed between 1986 and 2014 as part of the National Health Interview Surveys (NHIS) and who were linked to the National Death Index (NDI) through 2015. The NHIS includes annual cross-sectional household surveys administered by the National Center for Health Statistics (NCHS) that provide representative samples of the civilian noninstitutionalized U.S. population (NCHS 2015, 2018a). Restricted-use NHIS files with geographic data allowed for linking ambient pollution estimates at the census-tract level. Mortality follow-up information through 31 December 2015 was available from restricted-use NHIS files linked with the NDI, as described elsewhere (NCHS 2018b). The present analysis was based on two constructed NHIS cohorts, referred to as the full cohort and the subcohort. The full cohort consisted of 1,599,329 adults with available information for age, sex, race-ethnicity, income, marital status, educational attainment, census tract, estimated ambient pollution exposure, interview date, mortality status, and date of death (if deceased). Analyses were also performed on a subcohort of 635,539 respondents from the full cohort for whom body mass index (BMI) and smoking status data were also available.

Changes in the NHIS questionnaire across survey years (NCHS 2018a) necessitated harmonization of several key variables. Marital status, educational attainment, race, and Hispanic origin were reported relatively consistently during this time period, allowing simple changes to be made to create consistent categories. Because BMI was not directly available for those interviewed from 1986–1996, it was calculated using reported height and weight. The most substantive harmonization involved adjusting household income variables for inflation, using the Consumer Price Index with 2015 as a base year. Individuals whose income was reported as being within a range (e.g., \$35,000–50,000) were assigned the mean of the range (e.g., \$42,500), whereas those whose income was reported as being over a certain threshold (e.g., >\$75,000) were assigned the threshold value (e.g., \$75,000). These income values were adjusted for inflation, and uniform income categories were assigned to individuals based on their estimated inflation-adjusted household income.

Procedures for informed consent and data collection and linkage of the NHIS files were approved by the NCHS. The construction of analytic data files and the analyses of the restricted-use data were conducted in compliance with procedures that assured that subjects remained de-identified. Statistical analyses were conducted at the NCHS Research Data Center (RDC) in Hyattsville, MD, and research output was reviewed to ensure no disclosure risk to NHIS survey respondents. Because research reported in this manuscript uses publicly accessible data that are de-identified, it is not subject to federal regulations on protection of human research subjects.

### *Air Pollution Concentrations*

Nationwide regulatory monitoring for $PM_{2.5}$ started in 1999. Primary air pollution estimates employed here are annual-average $PM_{2.5}$ concentrations for 17 y (1999–2015), derived from regulatory monitoring data and constructed in a universal kriging framework; these were estimated by partial least squares from hundreds of geographic variables, including land use, population, and satellite-derived estimates of land use and air pollution. Hold-out cross-validation (CV) indicated good model performance (10-fold CV-$R^2$: 0.78–0.90). Detailed description, documentation, and evaluation of this modeling approach has been provided elsewhere (Kim et al. 2018). Modeled air pollution estimates for $PM_{2.5}$ (and other criteria pollutants) are freely accessible at the Center for Air, Climate, and Energy Solutions website (https://www.caces.us/).

In this study, annual $PM_{2.5}$ concentrations were estimated for 2000 and 2010 U.S. Census block centroids in the contiguous United States with nonzero population. Population-weighted annual averages were calculated for all 17 y for each 2000 and 2010 U.S. Census tract. Individuals were assigned air pollution concentrations based on their census tract of residency at the time of the survey, using year-2000 U.S. Census tracts for individuals surveyed from 1986 through 2010 and year-2010 U.S. Census tracts for those surveyed from 2011 through 2014. For primary analyses, $PM_{2.5}$ exposures were assigned to all cohort members as census-tract level average concentrations of $PM_{2.5}$ over the 17 y with regulatory monitoring and modeled $PM_{2.5}$ estimates (1999–2015).

To explore sensitivity to differences in exposure window and because the cohort included individuals who were surveyed prior to the 17-y (1999–2015) exposure window, census-tract level mean $PM_{2.5}$ concentrations were also estimated for a longer, 28-y (1988–2015) exposure window using back-casted, imputed $PM_{2.5}$ estimates for 1988–1998. Nationwide regulatory monitoring of $PM_{10}$ (particles <10 μm in aerodynamic diameter) began in 1988. Annual-average $PM_{10}$ concentration estimates for each census tract for 1988–2015 were modeled using the universal kriging modeling framework used to estimate $PM_{2.5}$ as noted above (Kim et al. 2018; https://www.caces.us/). Back-casted $PM_{2.5}$ concentrations were imputed at the census-tract level for each year from 1988 through 1998 by multiplying the census tract's mean $PM_{2.5}$:$PM_{10}$ ratio for 1999–2003 with that year's modeled $PM_{10}$ concentration (as illustrated in Figure S1). Mean concentrations of $PM_{2.5}$ over the 28-y exposure window (1988–2015) were estimated using imputed data from 1988–1998 and primary modeled data from 1999–2015. Although there was a downward trend in estimated $PM_{2.5}$ concentrations over time, the primary estimated 17-y (1999–2015) mean $PM_{2.5}$ concentrations were highly correlated with both imputed mean $PM_{2.5}$ concentrations from 1988–1998 ($r = 0.90$) and mean concentrations of $PM_{2.5}$ over the 28-y exposure window (1988–2015) ($r = 0.97$) in the full cohort.

### *Statistical Methods*

Adjusted hazard ratios (HRs) and 95% confidence intervals (CIs) relating relative morality risk associated with a 10-$\mu g/m^3$ increase of ambient $PM_{2.5}$ were estimated using predetermined Cox proportional hazards (CPH) regression models. Two variations of the CPH model were used. First, in order to account for NHIS's complex, stratified, multistage sample design, complex CPH models were estimated using eligibility-adjusted sample weights (to account for oversampling) from the 2015 NHIS Linked Mortality File (NCHS 2018b) with stratification by sampling strata and clustering by primary sampling units (to more accurately estimate standard errors) using the SURVEYPHREG procedure in SAS (version 9.3; SAS Institute Inc.). Second, basic CPH models, which did not account for complex survey design (using PHREG procedure in SAS version 9.3), were estimated as part of extended sensitivity and stratified analyses.

Survival times, in days, were calculated with date of interview treated as beginning of follow-up. End of follow-up was date of death for those who died; censored survival times, for survivors,

were the end of mortality follow-up (31 December 2015). $PM_{2.5}$–mortality HRs were estimated for different cause-of-death groupings. Deaths prior to 1999 were coded using the ninth revision of the International Statistical Classification of Diseases, Injuries, and Causes of Death (ICD-9); deaths from 1999 on were coded using the tenth revision (ICD-10) (NCHS 2018b). Deaths coded under ICD-9 guidelines were recoded into comparable ICD-10–based cause-of-death groups. Cause-of-death groupings used in this analysis include cardiopulmonary disease subdivided by cardiovascular disease (I00–I09, I11, I13, I20–I51), cerebrovascular disease (I60–I69), chronic lower respiratory disease (J40–J47), and influenza/pneumonia (J09–J18); cancers (C00–C97) with a focus on lung cancer (C33–C34); and all other or unknown causes of death. For analysis of specific cause-of-death groupings, censored survival times for the deceased were dates of death for any other cause of death.

$PM_{2.5}$ was included as a continuous variable in the CPH models. Models controlled for combinations of age, sex, and race-ethnicity. In basic CPH models, all 536 strata of 1-y age group, sex, and race-ethnicity were given their own baseline hazard (by including them in the STRATA statement of SAS, PHREG). Because of modeling and computational constraints, in complex CPH models, age, sex, and race-ethnicity were controlled for by including 104 variables indicating all interactive combinations of 13 age ranges (18–24 y, and each subsequent 5-y age groups), sex, and race-ethnicity as covariates in the model. Both complex and basic CPH models controlled for additional covariates by including indicator variables for levels of income inflation-adjusted to 2015 ($0–35,000; $35,000–50,000; $50,000–75,000; >$75,000); education levels (less than high school graduate, high school graduate, some college, college graduate, more than college graduate); marital status (married, divorced, separated, never married, widowed); rural versus urban, as defined by the U.S. Census Bureau (2018); U.S. census regions (Northeast, Midwest, South, West); and survey years (for each of the 29 survey years, 1986–2014). For analysis using the subcohort, indicator variables for smoking status (never, current, former) and BMI (<20, 20–25, 25–30, 30–35, $>35\,kg/m^2$) were also included in the models.

Sensitivity analyses were conducted by comparing $PM_{2.5}$–mortality HR estimates for complex and basic CPH models, analysis based on the full and subcohorts, and a series of models that, in a stepwise fashion, progressively added variables to the models. For the primary analyses, $PM_{2.5}$ exposures were assigned as the mean concentrations of $PM_{2.5}$ over the 17 y with regulatory monitoring and modeled $PM_{2.5}$ estimates (1999–2015). $PM_{2.5}$–mortality HRs were also estimated using a longer, 28-y exposure window (1988–2015) that included back-casted, imputed $PM_{2.5}$ based on $PM_{2.5}$:$PM_{10}$ ratios. In addition, $PM_{2.5}$-mortality HRs were estimated using the primary 17-y exposure period and restricting the analysis to only cohort members who were surveyed and followed up during this period. Subgroup analyses, using basic CPH models, were performed across three age groups (18–64, 65–75, >75 y at time of NHIS interview), sex, race-ethnicity, smoking status, BMI ranges, income level, marital status, urban-rural designation, census region, and 3 survey year groups.

The shape of the $PM_{2.5}$–mortality relationship was also explored using an integrated modeling approach that fit a class of flexible algebraic concentration–response functions, as documented elsewhere (Nasari et al. 2016). Briefly, a class of flexible functions was constructed by defining transformations of concentration as the product of either a linear or log-linear function of concentration multiplied by a logistic weighting function—allowing for flexible but monotonically nondecreasing concentration response functions. The estimation method was based on a routine that fit models within the class of concentration–response functions and selected the best fitting model.

## Results

A detailed summary of unweighted baseline characteristics for the full cohort of 1,599,329 subjects (267,204 deaths) and the subcohort of 635,539 subjects (106,385 deaths) is presented in Table 1.

**Table 1.** Baseline unweighted characteristics of the full and subcohorts created from U.S. National Health Interview Surveys from 1986–2014 with mortality follow-up through 2015.

| Variable | Full cohort | Subcohort |
|---|---|---|
| Total number in cohort | 1,599,329 | 635,539 |
| Total deaths | 267,204 | 106,385 |
| Cardiopulmonary[a] | 106,796 | 43,195 |
| Cardiovascular | 70,506 | 28,345 |
| Cerebrovascular | 15,502 | 6,297 |
| Chronic lower respiratory | 14,770 | 6,156 |
| Influenza/pneumonia | 6,018 | 2,397 |
| Cancers[a] | 68,202 | 26,453 |
| Lung cancers | 18,770 | 7,420 |
| All other or unknown | 92,206 | 36,737 |
| Sex (%) | | |
| Male | 47.15 | 44.54 |
| Female | 52.85 | 55.46 |
| Age [y (mean)] | 43.9 | 45.3 |
| Race-ethnicity (%) | | |
| Non-Hispanic white | 66.77 | 67.51 |
| Hispanic | 15.37 | 14.08 |
| Non-Hispanic black | 13.00 | 14.01 |
| All other or unknown | 4.86 | 4.40 |
| Income (inflation adjusted to 2015) (%) | | |
| $0–35,000 | 30.71 | 38.04 |
| $35,000–50,000 | 15.39 | 15.47 |
| $50,000–75,000 | 20.43 | 18.79 |
| >$75,000 | 33.47 | 27.71 |
| Marital Status (%) | | |
| Married | 60.72 | 49.57 |
| Divorced | 9.27 | 14.06 |
| Separated | 2.51 | 3.59 |
| Never married | 22.02 | 24.31 |
| Widowed | 5.48 | 8.47 |
| Education (%) | | |
| <High school graduate | 19.46 | 18.63 |
| High school graduate | 32.42 | 30.37 |
| Some college | 25.63 | 27.10 |
| College graduate | 14.06 | 15.03 |
| >College graduate | 8.42 | 8.87 |
| Urban/rural (%) | | |
| Urban | 76.57 | 77.64 |
| Rural | 23.43 | 22.36 |
| U.S. Census region (%) | | |
| Northeast | 18.81 | 18.08 |
| Midwest | 22.89 | 23.71 |
| South | 35.28 | 35.74 |
| West | 23.02 | 22.46 |
| BMI [$kg/m^2$ (%)] | | |
| <20 | — | 7.28 |
| 20–25 | — | 36.37 |
| 25–30 | — | 33.80 |
| 30–35 | — | 14.43 |
| >35 | — | 8.12 |
| Smoking (%) | | |
| Never | — | 53.76 |
| Current | — | 23.90 |
| Former | — | 22.34 |
| $PM_{2.5}$ (mean, SD, range)[b] | 10.7, 2.4, 2.5–19.2 | 10.7, 2.4, 2.5–19.2 |

Note: The data were complete for all variables listed for each cohort. For the subcohort, subjects also had complete data for BMI and smoking status. —, data not available; BMI, body mass index; $PM_{2.5}$, particulate matter <2.5 μm in aerodynamic diameter); SD, standard deviation.

[a]Cause-of-death groupings are based on International Statistical Classification of Diseases, Injuries, and Causes of Death, Tenth Revision (ICD-10) codes and include cardiopulmonary disease subdivided by cardiovascular disease (I00–I09, I11, I13, I20–I51), cerebrovascular disease (I60–I69), chronic lower respiratory disease (J40–J47), and influenza and pneumonia (J09–J18); cancers (C00–C97); and lung cancer (C33–C34).

[b]Mean modeled 1999–2015 $PM_{2.5}$ exposure for cohort subjects based on census tract of residence at time of survey.

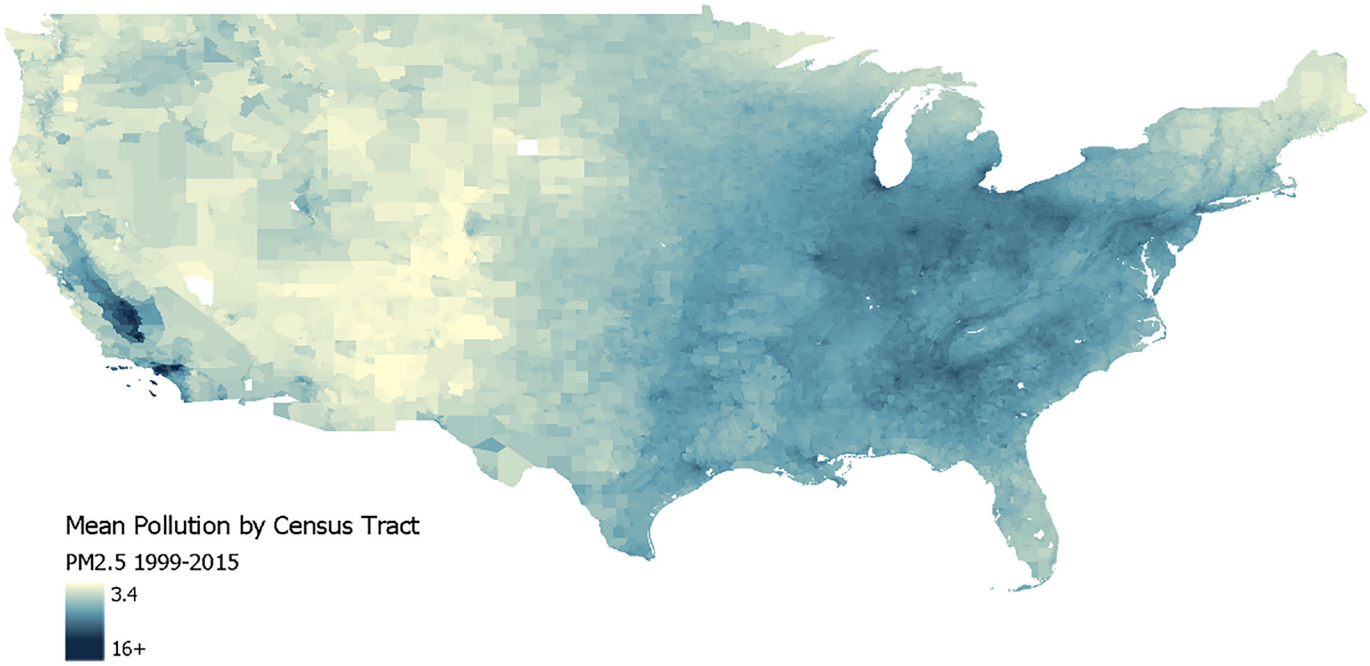


**Figure 1.** Average concentrations of $PM_{2.5}$ pollution ($\mu g/m^3$) by 2010 U.S. Census tracts in the continental United States, 1999–2015. $PM_{2.5}$, particulate matter <2.5 μm in aerodynamic diameter.

Figure 1 illustrates modeled $PM_{2.5}$ air pollution data for 1999–2015 at the 2010 census-tract level. Mean ambient 1999–2015 $PM_{2.5}$ concentrations for both the full and subcohort was 10.7 $\mu g/m^3$ with a range of 2.5 to 19.2 $\mu g/m^3$.

Estimated adjusted HRs (and 95% CIs) associated with a 10-$\mu g/m^3$ increase of $PM_{2.5}$ for various causes of death using both the full and subcohorts and using the full complex CPH model are provided in Table 2. Elevated long-term exposure to $PM_{2.5}$ was associated with elevated risk of all-cause, cardiopulmonary, cardiovascular, cerebrovascular, influenza pneumonia, cancer, and lung cancer mortality. Estimated HRs for key covariables (for all-cause, cardiopulmonary, and lung cancer mortality) in the model using the subcohort and the complex CPH model are presented in Table S1. In addition to $PM_{2.5}$ exposure and, as expected, higher mortality risks were associated with lower income, marital status other than married, lower education, being underweight or obese, and smoking (see Table S1).

**Table 2.** Estimated hazard ratios (95% CIs) associated with 10 $\mu g/m^3$ $PM_{2.5}$ for different causes of death using both the full and subcohorts using the complex CPH model.

| | Full cohort | | Subcohort | |
|---|---|---|---|---|
| Cause of death[a] | Hazard ratio | 95% CI | Hazard ratio | 95% CI |
| All cause | 1.13 | 1.11, 1.16 | 1.12 | 1.08, 1.15 |
| Cardiopulmonary | 1.24 | 1.20, 1.29 | 1.23 | 1.17, 1.29 |
| Cardiovascular | 1.30 | 1.25, 1.36 | 1.28 | 1.21, 1.36 |
| Cerebrovascular | 1.27 | 1.16, 1.39 | 1.26 | 1.11, 1.43 |
| Chronic lower respiratory | 0.93 | 0.85, 1.03 | 0.95 | 0.82, 1.09 |
| Influenza/pneumonia | 1.47 | 1.27, 1.71 | 1.41 | 1.13, 1.76 |
| Cancers | 1.14 | 1.09, 1.19 | 1.15 | 1.08, 1.22 |
| Lung cancer | 1.08 | 0.99, 1.18 | 1.12 | 1.00, 1.26 |
| Other or unknown cause | 1.02 | 0.99, 1.06 | 0.99 | 0.94, 1.05 |

Note: Adjusted for age, sex, race, income, education, marital status, urban versus rural, census regions, survey year and the complex NHIS survey design. The subcohort was additionally adjusted for smoking status and BMI. BMI, body mass index; CI, confidence interval; CPH, Cox proportional hazards (regression model); NHIS, National Health Interview Surveys; $PM_{2.5}$, particulate matter <2.5 μm in aerodynamic diameter).
[a]Cause-of-death groupings are based on International Statistical Classification of Diseases, Injuries, and Causes of Death, Tenth Revision (ICD-10) codes and include cardiopulmonary disease subdivided by cardiovascular disease (I00–I09, I11, I13, I20–I51), cerebrovascular disease (I60–I69), chronic lower respiratory disease (J40–J47), and influenza and pneumonia (J09–J18); cancers (C00–C97); and lung cancer (C33–C34).

Figure 2 illustrates the model sensitivity analysis (corresponding numeric data for these results are presented in Table S2). Estimated $PM_{2.5}$–mortality HRs were not sensitive to cohort selection or modeling choices. Estimated $PM_{2.5}$–mortality HRs were nearly the same for the full and subcohort and when using complex versus basic CPH models. The 95% CIs were slightly wider for the complex CPH models. After controlling for combinations of age, sex, and race-ethnicity, there was only marginal attenuation of the $PM_{2.5}$–mortality HRs with additional covariates in the models. Controlling for rural versus urban or U.S. census region had minimal impact on the estimated $PM_{2.5}$–mortality HRs. Controlling for survey year by including 29 survey year indicator variables in the model also had minimal impact on the estimated $PM_{2.5}$–mortality HRs. The estimated $PM_{2.5}$–mortality HRs were somewhat attenuated when a 28-y exposure window (1988–2015) using back-casted, imputed $PM_{2.5}$ (for 1988–1998) was used. The estimated $PM_{2.5}$–mortality HRs, however, were larger when using the primary 17-y exposure period and restricting the analysis to cohort members who were surveyed during this period.

Figure 3 illustrates the results from the stratified or subgroup analyses using the basic CPH model (corresponding numeric data for these results are presented in Table S3). In each stratification, the models controlled for all other covariates. Results were nearly identical for males and females, and there were no consistent or coherent differences across strata for race-ethnicity, BMI, income, marital status, urban-rural designation, or survey year. For all-cause mortality, the HR was larger for relatively younger subjects (<65 y of age at the time of the survey). HRs were generally larger for never-smokers (especially for lung cancer). Estimated $PM_{2.5}$–mortality associations for all-cause mortality were positive for all census regions, with the largest HRs in the Midwest.

The shape of the estimated concentration–response relationship between $PM_{2.5}$ and all-cause and cardiopulmonary mortality,

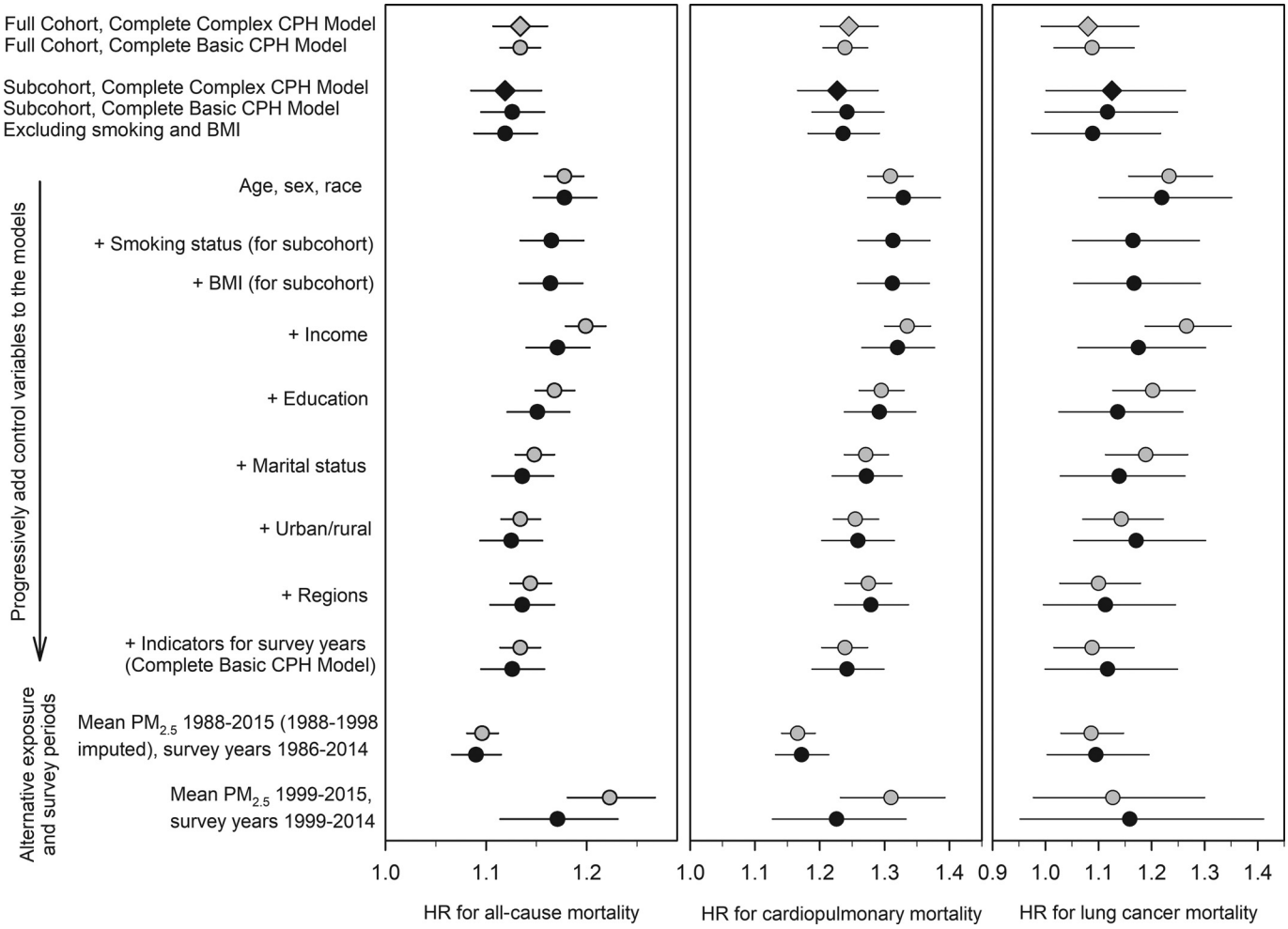


**Figure 2.** Illustration of model sensitivity analysis. Hazard ratios (and 95% CIs) associated with 10 $\mu g/m^3$ $PM_{2.5}$ estimated from various models are presented. Gray and black symbols indicate models that use the full cohort and subcohort, respectively. Diamonds indicate complex CPH models that control for the complex survey design, whereas circles indicate models that use the basic CPH models. Cause-of-death groupings are based on ICD-10 codes. Cardiopulmonary disease includes cardiovascular disease (I00–I09, I11, I13, I20–I51), cerebrovascular disease (I60–I69), chronic lower respiratory disease (J40–J47), and influenza and pneumonia (J09–J18). Lung cancer includes C33–C34. CI, confidence interval; CPH, Cox proportional hazards (regression model); HR, hazard ratio; $PM_{2.5}$, particulate matter <2.5 μm in aerodynamic diameter; ICD-10, International Statistical Classification of Diseases, Injuries, and Causes of Death, Tenth Revision.

using the subcohort, controlling for all covariates including smoking and BMI, and using the flexible modeling approach (Nasari et al. 2016) is illustrated in Figure 4. For all-cause mortality there was some evidence of a flatter response relationship at the lower concentrations (less than $\sim 8\ \mu g/m^3$). For cardiopulmonary mortality, the fit was nearly linear.

## Discussion

This study observed, in a large, representative, contemporary cohort of U.S. adults, that long-term exposure to $PM_{2.5}$ air pollution was associated with elevated risks of early mortality. Statistical models used to estimate $PM_{2.5}$–mortality HRs were predetermined and *a priori* model results are fully presented. Estimated $PM_{2.5}$–mortality HRs were statistically robust and not highly sensitive to key modeling choices. The increased mortality risk was primarily associated with cardiopulmonary mortality, including cardiovascular, cerebrovascular, and influenza/pneumonia. $PM_{2.5}$ air pollution was also associated with lung cancer mortality in never-smokers. With regard to respiratory disease, $PM_{2.5}$ air pollution was associated with influenza/pneumonia but not chronic lower respiratory disease, a finding that was also observed in the U.S. ACS CPS-II cohort (Pope et al. 2004).

Overall results of this study are comparable to other key studies. For example, in this analysis a 10-$\mu g/m^3$ long-term elevation in $PM_{2.5}$ was associated with an estimated all-cause mortality HR of approximately 1.12 (95% CI: 1.08, 1.15). The estimated HR from this NHIS cohort is smaller than the HR estimates observed in the Six Cities cohort (Dockery et al. 1993; Lepeule et al. 2012) (1.14; 95% CI: 1.07, 1.22) or the 2001 Canadian Census Health and Environment Cohort (Pinault et al. 2017) (1.15; 95% CI: 1.12, 1.17) but it is larger than estimates from the ACS CPS-II cohort (Pope et al. 2002, 2015; Jerrett et al. 2017) (1.07; 95% CI: 1.06, 1.09) and the U.S. Medicare cohort (Di et al. 2017) (1.073; 95% CI: 1.071, 1.075).

Two previous efforts to evaluate mortality effects of $PM_{2.5}$ air pollution using NHIS data have been made (Parker et al. 2018; Pope et al. 2018). The first (Parker et al. 2018) used 1997–2009 NHIS data with mortality follow-up through 2011 with a focus on evaluating effects of $PM_{2.5}$ on heart disease mortality by race and ethnicity. The second (Pope et al. 2018) used public-use NHIS data to conduct the analysis using fully publicly available data and in preparation for the current analysis at the NCHS RDC using the more extensive limited-use data. The analysis was limited to the 1986–2001 NHIS survey years with mortality follow-up through 2011. Restricted geographic information available in

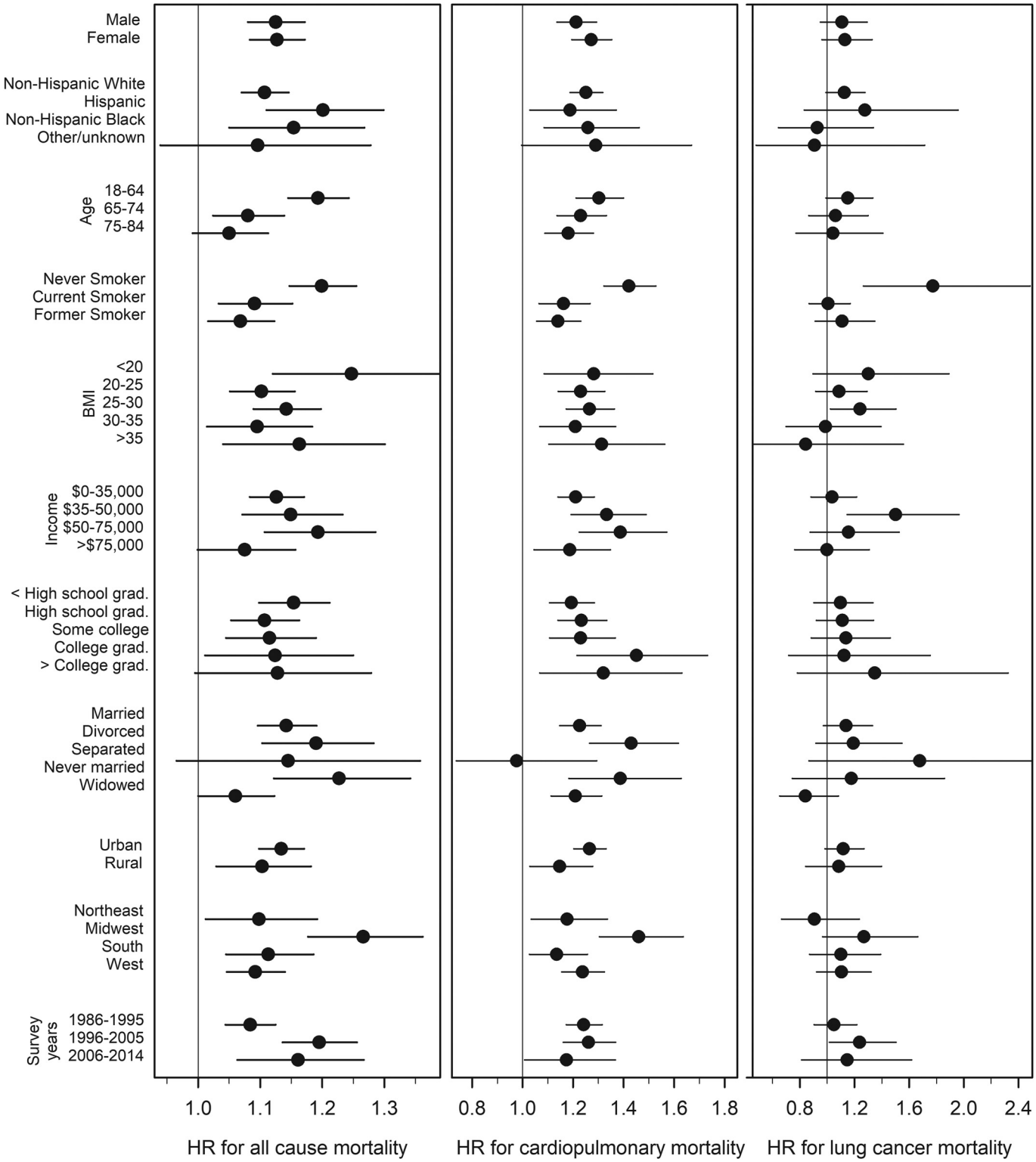


**Figure 3.** Illustration of stratified analysis for the subcohort. Hazard ratios (and 95% CIs) associated with 10 $\mu g/m^3$ $PM_{2.5}$ estimated from the basic CPH model are presented by sex, race-ethnicity, age, smoking status, BMI, income, education, marital status, rural/urban, census regions, and survey years. All stratified estimates are adjusted for remaining covariates. Cause-of-death groupings are based on ICD-10 codes. Cardiopulmonary disease includes cardiovascular disease (I00–I09, I11, I13, I20–I51), cerebrovascular disease (I60–I69), chronic lower respiratory disease (J40–J47), and influenza and pneumonia (J09–J18). Lung cancer includes C33–C34. BMI, body mass index; CI, confidence interval; CPH, Cox proportional hazards (regression model); HR, hazard ratio; $PM_{2.5}$, particulate matter <2.5 μm in aerodynamic diameter; ICD-10, International Statistical Classification of Diseases, Injuries, and Causes of Death, Tenth Revision.

the public-use data allowed only for inclusion of individuals who resided in large metropolitan areas, and $PM_{2.5}$ exposures could only be assigned at the metropolitan statistical area level. Both studies observed positive $PM_{2.5}$–mortality associations for all-cause and cardiovascular disease mortality, but with less statistical precision. For example, a 10-$\mu g/m^3$ long-term elevation in

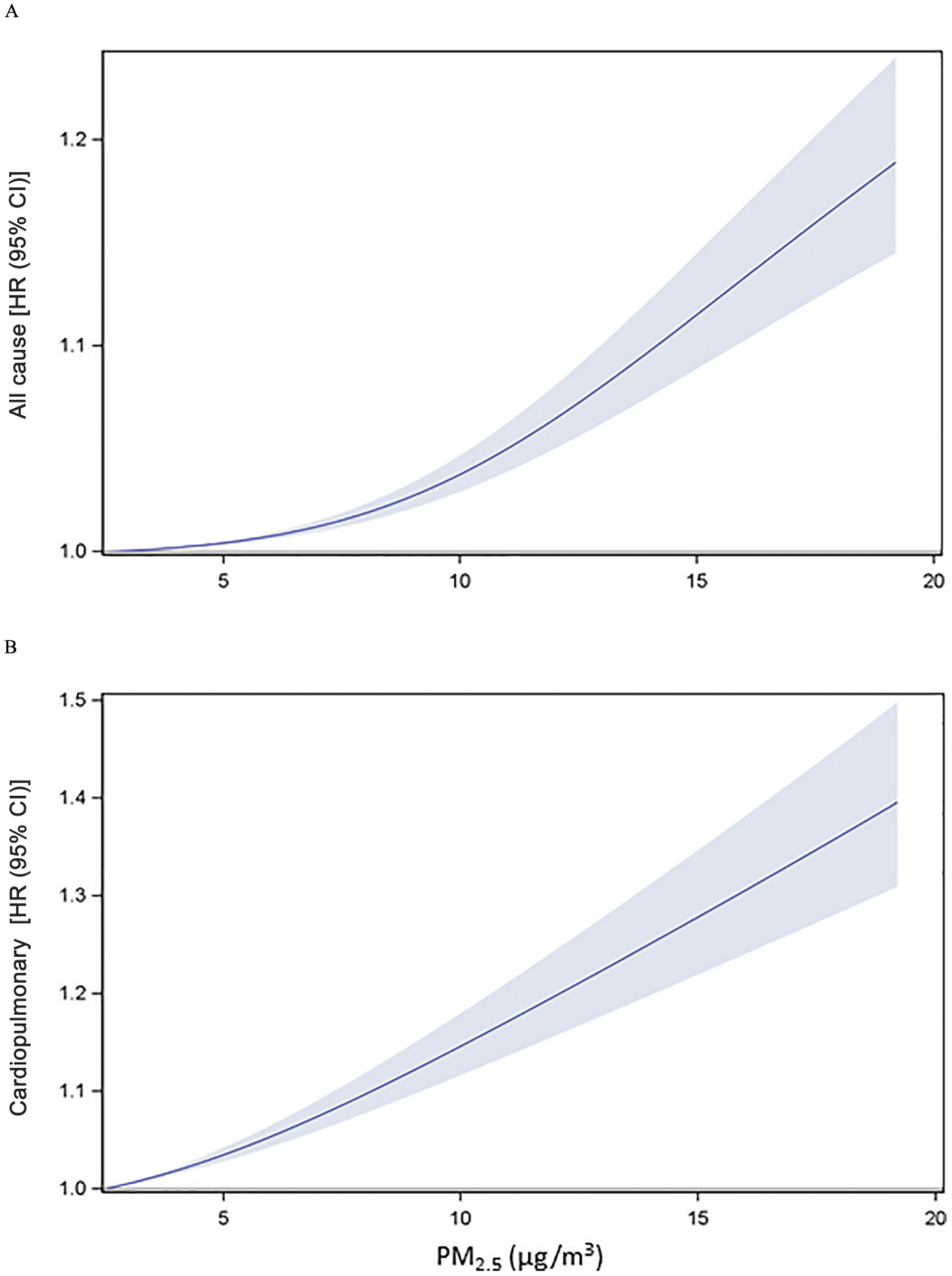


**Figure 4.** Estimated concentration–response associations between $PM_{2.5}$ and all-cause (A) and cardiopulmonary (B) mortality using the subcohort and basic CPH model with the flexible modeling approach, adjusting for age, sex, race-ethnicity, income, education, marital status, urban versus rural, census regions, survey year, smoking status, and BMI. The optimal nonlinear models are presented as solid lines with 95% uncertainty bounds (shaded area). Cause-of-death groupings are based on ICD-10 codes. Cardiopulmonary disease includes cardiovascular disease (I00–I09, I11, I13, I20–I51), cerebrovascular disease (I60–I69), chronic lower respiratory disease (J40–J47), and influenza and pneumonia (J09–J18). BMI, body mass index; CI, confidence interval; CPH, Cox proportional hazards (regression model); HR, hazard ratio; ICD-10, International Statistical Classification of Diseases, Injuries, and Causes of Death, Tenth Revision; $PM_{2.5}$, particulate matter <2.5 μm in aerodynamic diameter.

$PM_{2.5}$ was associated with an estimated all-cause mortality HR of approximately 1.08 (95% CI: 1.01, 1.16) and 1.06 (95% CI: 1.01, 1.11) for the two studies, respectively.

This study has several important strengths: *a*) It was based on representative samples of U.S. adults with high-quality and well-documented survey design and methods, survey interviews, and data quality management. Survey respondents represented a range of values for demographics (e.g., age, sex, race-ethnicity, income, education) and geographies (urban/rural, U.S. Census region). *b*) This constructed NHIS cohort was large, providing substantial statistical power. *c*) $PM_{2.5}$–mortality HRs could be estimated for all-cause mortality and for mortality from various relevant cause-of-death groupings. *d*) The analysis could control for key individual risk factors, including smoking status in the subcohort. Interestingly, with

control for age, sex, race-ethnicity, income, education, marital status, BMI, rural versus urban, census regions, and survey years, the estimated $PM_{2.5}$–mortality HRs were not sensitive to the inclusion of smoking status in the models. This finding is suggestive that analyses of administrative cohorts without smoking data, such as the U.S. Medicare cohort (Di et al. 2017) or the Canadian Census Health and Environment Cohort (Pinault et al. 2017), can be informative. *e*) The positive $PM_{2.5}$–mortality associations for all-cause and cardiopulmonary mortality were not highly sensitive to cohort selection or modeling choices. *f*) Ambient air pollution estimates are publicly available at the census-tract level throughout the continental United States, including both urban and rural areas. *g*) The NHIS files, with mortality follow-up and geographic information that allows for linking with air pollution data, are generally available for research purposes; NHIS manages the limited-use files to ensure no disclosure risk to survey respondents.

A primary limitation of this study, as in all air pollution studies, is the lack of direct measures of individual lifetime pollution exposures. Long-term $PM_{2.5}$ exposures must be estimated using available ground-based monitoring of ambient concentrations, land use regression, and related modeling. Furthermore, although the study cohort used survey data from 1986–2014 (29 y) and mortality follow-up from 1986–2015 (30 y), and although cohort subjects were exposed to air pollution in their lives prior to being surveyed, monitoring networks for $PM_{2.5}$ did not exist prior to 1999. Therefore, modeled air pollution data are only available for 1999–2015 (17 y). Back-casted, imputed $PM_{2.5}$ for 1988–1998 indicate that $PM_{2.5}$ exposures were higher prior to 1999, but the imputed pre-1999 $PM_{2.5}$ concentrations were highly correlated with the $PM_{2.5}$ concentrations from 1999–2015—suggesting that the 17-y mean $PM_{2.5}$ concentrations are partially indicative of longer-term exposures. The generally declining pollution levels and the high spatial correlation across the time periods could result in scaling bias, but the direction of that bias is largely dependent upon the relevant exposure window. There was some sensitivity to the $PM_{2.5}$–mortality association when the exposure window was extended with back-casted $PM_{2.5}$ data or when the analysis was restricted to subjects followed up only during the 17-y period with reliable modeled $PM_{2.5}$ data—yet positive and somewhat comparable associations were still observed. In addition, exposure assignment does not account for subjects moving during the follow-up period. It may be presumed that exposure measurement error would likely bias the $PM_{2.5}$–mortality estimates to the null, but the potential for higher $PM_{2.5}$ concentrations prior to 1999 and the potential of compression of the exposure distribution, as a result of exposure modeling, make it difficult to determine the overall direction of bias.

Another limitation of this—and all observational studies—is the potential of residual confounding because of some unknown, unmeasured, or inadequately controlled-for risk factor that is associated with mortality while also correlated with ambient air pollution exposures. This analysis controlled for age, sex, and race-ethnicity along with various key individual risk factors and other factors such as urban/rural, geographic regions, and survey years. Although the estimated $PM_{2.5}$–mortality HRs were reasonably consistent across modeling choices and covariates included in the models, there remains the possibility of residual confounding.

In conclusion, this study substantially expands the evidence that long-term exposure to fine particulate matter air pollution contributes to risk of mortality—especially cardiopulmonary and lung cancer mortality. These results are uniquely based on a large, nationwide, representative cohort of U.S. adults. $PM_{2.5}$–mortality associations are observed widely across subgroups of sex, age, race, ethnicity, income and education levels, and broad geographic regions. The estimated excess risks from exposure to $PM_{2.5}$ air pollution to any given individual are certainly not as large as several other individual risk factors such as cigarette smoking, poverty, or obesity (see Table S1). However, given the ubiquitous and involuntary nature of $PM_{2.5}$ exposures and given the impact on burden of disease (Cohen et al. 2017), these results are of substantial public health importance.

## Acknowledgments

This publication was developed as part of the Center for Air, Climate, and Energy Solutions (CACES), which was supported under Assistance Agreement No. R835873 awarded by the U.S. Environmental Protection Agency (EPA). It has not been formally reviewed by the U.S. EPA. The views expressed in this document are solely those of authors and do not necessarily reflect those of the agency. The U.S. EPA does not endorse any products or commercial services mentioned in this publication. Findings and conclusions of this research are those of the authors and do not necessarily represent the views of the National Center for Health Statistics (NCHS) Research Data Center (RDC), the NCHS, the U.S. EPA, the Centers for Disease Control and Prevention, or Cornerstone Research.

## References

Beelen R, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, Hoffmann B, 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(9919):785–795, PMID: 24332274, https://doi.org/10.1016/S0140-6736(13)62158-3.

Bentayeb M, Wagner V, Stempfelet M, Zins M, Goldberg M, Pascal M, et al. 2015. Association between long-term exposure to air pollution and mortality in France: a 25-year follow-up study. Environ Int 85:5–14, PMID: 26298834, https://doi.org/10.1016/j.envint.2015.08.006.

Brook RD, Rajagopalan S, Pope CA III, Brook JR, Bhatnagar A, Diez-Roux AV, et al. 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. Circulation 121(21):2331–2378, PMID: 20458016, https://doi.org/10.1161/CIR.0b013e3181dbece1.

Carey IM, Atkinson RW, Kent AJ, van Staa T, Cook DG, Anderson HR. 2013. Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. Am J Respir Crit Care Med 187(11):1226–1233, PMID: 23590261, https://doi.org/10.1164/rccm.201210-1758OC.

Cesaroni G, Badaloni C, Gariazzo C, Stafoggia M, Sozzi R, Davoli M, et al. 2013. Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. Environ Health Perspect 121(3):324–331, PMID: 23308401, https://doi.org/10.1289/ehp.1205862.

Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, et al. 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. Lancet 389(10082):1907–1918, PMID: 28408086, https://doi.org/10.1016/S0140-6736(17)30505-6.

Crouse DL, Peters PA, Hystad P, Brook JR, van Donkelaar A, Martin RV, et al. 2015. Ambient $PM_{2.5}$, $O_3$, and $NO_2$ exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). Environ Health Perspect 123(11):1180–1186, PMID: 26528712, https://doi.org/10.1289/ehp.1409276.

Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. 2017. Air pollution and mortality in the Medicare population. N Engl J Med 376(26):2513–2522, PMID: 28657878, https://doi.org/10.1056/NEJMoa1702747.

Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, et al. 1993. An association between air pollution and mortality in six U.S. cities. N Engl J Med 329(24):1753–1759, PMID: 8179653, https://doi.org/10.1056/NEJM199312093292401.

Fischer PH, Marra M, Ameling CB, Hoek G, Beelen R, de Hoogh K, et al. 2015. Air pollution and mortality in seven million adults: the Dutch Environmental Longitudinal Study (DUELS). Environ Health Perspect 123(7):697–704, PMID: 25760672, https://doi.org/10.1289/ehp.1408254.

GBD 2015 Risk Factors Collaborators. 2016. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 388:1659–1724, PMID: 27733284, https://doi.org/10.1016/S0140-6736(16)31679-8.

Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, et al. 2014. Outdoor particulate matter exposure and lung cancer: a systematic review and

meta-analysis. Environ Health Perspect 122(9):906–911, PMID: 24911630, https://doi.org/10.1289/ehp/1408092.

Hart JE, Liao X, Hong B, Puett RC, Yanosky JD, Suh H, et al. 2015. The association of long-term exposure to $PM_{2.5}$ on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. Environ Health 14:38, PMID: 25926123, https://doi.org/10.1186/s12940-015-0027-6.

Jerrett M, Turner MC, Beckerman BS, Pope CA III, van Donkelaar A, Martin RV, et al. 2017. Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates. Environ Health Perspect 125(4):552–559, PMID: 27611476, https://doi.org/10.1289/EHP575.

Kim S-Y, Bechle M, Hankey S, Sheppard EA, Szpiro AA, Marshall JD. 2018. Concentrations of criteria pollutants in the contiguous U.S., 1979–2015: role of model parsimony in integrated empirical geographic regression. (November 2018). UW Biostatistics Working Paper Series. Working Paper 425. https://biostats.bepress.com/uwbiostat/paper425 [accessed 28 May 2019].

Lepeule J, Laden F, Dockery D, Schwartz J. 2012. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. Environ Health Perspect 120(7):965–970, PMID: 22456598, https://doi.org/10.1289/ehp.1104660.

Lipsett MJ, Ostro BD, Reynolds P, Goldberg D, Hertz A, Jerrett M, et al. 2011. Long-term exposure to air pollution and cardiorespiratory disease in the California Teachers Study cohort. Am J Respir Crit Care Med 184(7):828–835, PMID: 21700913, https://doi.org/10.1164/rccm.201012-2082OC.

Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, et al. 2007. Long-term exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med 356(5):447–458, PMID: 17267905, https://doi.org/10.1056/NEJMoa054409.

Nasari MM, Szyszkowicz M, Chen H, Crouse D, Turner MC, Jerrett M, et al. 2016. A class of non-linear exposure-response models suitable for health impact assessment applicable to large cohort studies of ambient air pollution. Air Qual Atmos Health 9(8):961–972, PMID: 27867428, https://doi.org/10.1007/s11869-016-0398-z.

National Center for Health Statistics (NCHS). 2015. *2014 National Health Interview Survey: Survey Description*. Hyattsville, Maryland:NHIS, Division of Health Interview Statistics. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2014/srvydesc.pdf [accessed 31 August 2018].

NCHS. 2018a. National Health Interview Survey, 1986–2014. NHIS data, questionnaires and related documentation. https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm [accessed 31 August 2018].

NCHS. 2018b. NCHS data linked to NDI mortality files. https://www.cdc.gov/nchs/data-linkage/mortality.htm [accessed 31 August 2018].

Parker JD, Kravets N, Vaidyanathan A. 2018. Particulate matter air pollution exposure and heart disease mortality risks by race and ethnicity in the United States: 1997 to 2009 National Health Interview Survey with mortality follow-up through 2011. Circulation 137(16):1688–1697, PMID: 29237717, https://doi.org/10.1161/CIRCULATIONAHA.117.029376.

Pinault L, Tjepkema M, Crouse DL, Weichenthal S, van Donkelaar A, Martin RV, et al. 2016. Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian Community Health Survey cohort. Environ Health 15:18, PMID: 26864652, https://doi.org/10.1186/s12940-016-0111-6.

Pinault LL, Weichenthal S, Crouse DL, Brauer M, Erickson A, van Donkelaar AV, et al. 2017. Associations between fine particulate matter and mortality in the 2001 Canadian Census Health and Environment Cohort. Environ Res 159:406–415, PMID: 28850858, https://doi.org/10.1016/j.envres.2017.08.037.

Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287(9):1132–1141, PMID: 11879110, https://doi.org/10.1001/jama.287.9.1132.

Pope CA III, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D, et al. 2004. Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. Circulation 109(1):71–77, PMID: 14676145, https://doi.org/10.1161/01.CIR.0000108927.80044.7F.

Pope CA III, Ezzati M, Cannon JB, Allen RT, Jerrett M, Burnett RT. 2018. Mortality risk and $PM_{2.5}$ air pollution in the USA: an analysis of a national prospective cohort. Air Qual Atmos Health 11(3):245–252, https://doi.org/10.1007/s11869-017-0535-3.

Pope CA III, Turner MC, Burnett RT, Jerrett M, Gapstur SM, Diver WR, et al. 2015. Relationships between fine particulate air pollution, cardiometabolic disorders and cardiovascular mortality. Circ Res 116(1):108–115, PMID: 25348167, https://doi.org/10.1161/CIRCRESAHA.116.305060.

Puett RC, Hart JE, Suh H, Mittleman M, Laden F. 2011. Particulate matter exposures, mortality, and cardiovascular disease in the Health Professionals Follow-Up Study. Environ Health Perspect 119(8):1130–1135, PMID: 21454146, https://doi.org/10.1289/ehp.1002921.

Thurston GD, Ahn J, Cromar KR, Shao Y, Reynolds HR, Jerrett M, et al. 2016. Ambient particulate matter air pollution exposure and mortality in the NIH-AARP Diet and Health cohort. Environ Health Perspect 124(4):484–490, PMID: 26370657, https://doi.org/10.1289/ehp.1509676.

Tseng E, Ho W-C, Lin M-H, Cheng T-J, Chen P-C, Lin H-H. 2015. Chronic exposure to particulate matter and risk of cadiovascular mortality: cohort study from Taiwan. BMC Public Health 15:936–944, PMID: 26392179, https://doi.org/10.1186/s12889-015-2272-6.

U.S. Census Bureau. 2018. Rural America: How does the U.S. Census Bureau define "Rural" Interactive Story Map. https://gis-portal.data.census.gov/arcgis/apps/MapSeries/index.html?appid=7a41374f6b03456e9d138cb014711e01 [accessed 31 August 2018].

Villeneuve PJ, Weichenthal SA, Crouse D, Miller AB, To T, Martin RV, et al. 2015. Long-term exposure to fine particulate matter air pollution and mortality among Canadian women. Epidemiology 26(4):536–545, PMID: 25894856, https://doi.org/10.1097/EDE.0000000000000294.

Vodonos A, Awad YA, Schwartz J. 2018. The concentration-response between long-term $PM_{2.5}$ exposure and mortality; a meta-regression approach. Environ Res 166:677–689, PMID: 30077140, https://doi.org/10.1016/j.envres.2018.06.021.

Woodruff TJ, Parker JD, Schoendorf KC. 2006. Fine particulate matter ($PM_{2.5}$) air pollution and selected causes of postneonatal infant mortality in California. Environ Health Perspect 114(5):786–790, PMID: 16675438, https://doi.org/10.1289/ehp.8484.

Yin P, Brauer M, Cohen A, Burnett RT, Liu J, Liu Y, et al. 2017. Long-term fine particulate matter exposure and nonaccidental and cause-specific mortality in a large national cohort of Chinese men. Environ Health Perspect 125(11):117002, PMID: 29116930, https://doi.org/10.1289/EHP1673.