Original Research Article

ENVIRONMENTAL EPIDEMIOLOGY

OPEN

Estimating long-term pollution exposure effects through inverse probability weighting methods with Cox proportional hazards models

Joshua D. Higbee[a,*], Jacob S. Lefler[b], Richard T. Burnett[c], Majid Ezzati[d], Julian D. Marshall[e], Sun-Young Kim[f], Matthew Bechle[e], Allen L. Robinson[g], C. Arden Pope III[h]

Background: Fine particulate matter ($PM_{2.5}$) is associated with negative health outcomes in both the short and long term. However, the cohort studies that have produced many of the estimates of long-term exposure associations may fail to account for selection bias in pollution exposure as well as covariate imbalance in the study population; therefore, causal modeling techniques may be beneficial.

Methods: Twenty-nine years of data from the National Health Interview Survey (NHIS) was compiled and linked to modeled annual average outdoor $PM_{2.5}$ concentration and restricted-use mortality data. A series of Cox proportional hazards models, adjusted using inverse probability weights, yielded causal risk estimates of long-term exposure to ambient $PM_{2.5}$ on all-cause and cardiopulmonary mortality.

Results: Covariate-adjusted estimated relative risks per 10 μg/m^3 increase in $PM_{2.5}$ exposure were estimated to be 1.117 (1.083, 1.152) for all-cause mortality and 1.232 (1.174, 1.292) for cardiopulmonary mortality. Inverse probability weighted Cox models provide relatively consistent and robust estimates similar to those in the unweighted baseline multivariate Cox model, though they have marginally lower point estimates and higher standard errors.

Conclusions: These results provide evidence that long-term exposure to $PM_{2.5}$ contributes to increased mortality risk in US adults and that the estimated effects are generally robust to modeling choices. The size and robustness of estimated associations highlight the importance of clean air as a matter of public health. Estimated confounding due to measured covariates appears minimal in the NHIS cohort, and various distributional assumptions have little bearing on the magnitude or standard errors of estimated causal associations.

Key Words: Fine particulate matter; Inverse probability of treatment weighting; Mortality; National Health Interview Survey; Pollution

Introduction

The association between long-term exposure to fine particulate matter ($PM_{2.5}$, or particles less than 2.5 μm in aerodynamic diameter) and all-cause and specific cause mortality has been the subject of intensive research. $PM_{2.5}$ concentration in the atmosphere results in part from the use of coal, gasoline, and biofuels; the widespread use of these materials means that negative associations between pollution exposure and mortality risk have serious implications for public health. Numerous cohort studies have analyzed $PM_{2.5}$–mortality associations in careful detail with both representative (constructed to reflect a country or region's demographic characteristics) and nonrepresentative

What this study adds

This analysis is among the first to employ inverse probability weights in studying a continuous measure of fine particulate matter ($PM_{2.5}$) exposure and the first to do so using data from US National Health Interview Surveys. It also employs multiple distributions for more flexibility in computing these weights. The main findings of this study are statistically significant, causal effect estimates of long-term $PM_{2.5}$ exposure on all-cause and cardiopulmonary mortality. These estimates closely mirror the estimates yielded in prospective cohort studies using standard Cox proportional hazards models. These results are important and will be of interest to the readership of *Environmental Epidemiology*.

[a]Department of Economics, University of Chicago, Chicago, Illinois; [b]Department of Agricultural and Resource Economics, University of California – Berkeley, Berkeley, California; [c]Health Canada, Ottawa, Ontario, Canada; [d]MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom; [e]Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington; [f]Department of Cancer Control and Population Health, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang-si, Gyeonggi-do, Korea; [g]Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania; and [h]Department of Economics, Brigham Young University, Provo, Utah.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of the article.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.environepidem.com).

The results reported herein correspond to specific aims of grant R835873 to the Center for Air, Climate, and Energy Solutions (CACES) from the US Environmental Protection Agency (EPA).

Code and Data for Replication: Contact Joshua D. Higbee for all code used for this project. Air pollution estimates are available at the Center for Air, Climate, and Energy Solutions (CACES) website, while public-use National Health Interview Survey (NHIS) data are available at the National Center for Health Statistics (NCHS) website. Restricted-use mortality and geographic data may be obtained, subject to approval, from the NCHS.

All views and opinions expressed in this article are of the authors, and do not represent those of the Environmental Protection Agency (EPA), National Center for Health Statistics (NCHS), or the authors' institutions.

*Corresponding author. Address: Joshua D. Higbee, University of Chicago, 5757 South University Avenue, Chicago, IL 60637. E-mail address: joshuadhigbee@uchicago.edu (J.D. Higbee).

Environmental Epidemiology (2020) 4:e085

Received: 20 August 2019; Accepted 19 January 2020

Published online 14 February 2020

DOI: 10.1097/EE9.0000000000000085

(constructed to reflect a target group within the larger region) cohorts in North America,[1–17] Europe,[18–22] and Asia.[23,24] The results of these studies indicate associations between mortality risk and higher long-term exposure to $PM_{2.5}$, underscoring the importance of an accurate understanding of health considerations related to exposure to ambient air pollution.

The $PM_{2.5}$–mortality associations reported in the literature almost exclusively originate from studies using cohorts that were not constructed to study air pollution and are thus susceptible to potential bias through both selection bias and confounding across exposure levels, even after controlling for key covariates in the regression model itself. These studies include cohorts composed of subsets of the population where such confounding is intuitively likely, as well as selection bias due to nonrandom selection of study participants based on their belonging to particular subsets of the population.[1–6,8,9,13,15] Even nationally representative data sources used in other cohort studies may be affected by these issues—an increased probability of greater exposure may be associated with other covariates affecting survival, biasing the results through measured confounding. Additionally, cohorts constructed to be nationally representative may adequately represent distributions of key demographic characteristics without representing the national distribution of $PM_{2.5}$ concentrations depending on the locations from which study participants were sampled. Thus, the associations reported in numerous studies may be biased in either direction due to the potential correlations of measured exposure and other covariates.

While numerous prospective cohort studies have attempted to estimate the association between long-term pollution exposure and mortality risk, several other studies—including some using cohorts—have employed causal modeling techniques. While not strictly necessary in estimating a causal association, causal inference approaches provide additional causal evidence regarding the observed associations. One recent study employed a regression discontinuity design based on a Chinese policy of providing free or subsidized coal for indoor heating to areas north of the Huai river.[25] Wang et al[26] introduced a doubly robust additive hazards model that allows for the estimation of causal effects with a continuous pollution exposure measure through controlling for covariate imbalance across exposure levels in a cohort of Medicare beneficiaries, and Wu et al[27] used a similar estimator with the purpose of controlling for both exposure measurement error and covariate imbalance. Wang et al[28] also used a difference-in-differences approach to study exposure effects on a population in New Jersey, while Kioumourtzoglou et al[29] employed a similar design to examine trends in mortality within cities. Some recent cohort studies have examined pollution–mortality relationships with marginal structural models and inverse probability weighted logistic regressions.[30,31] Another analysis of health effects of pollution within a cohort was limited to binary cases, in which exposure is discretized based on being above or below a certain benchmark such as 12 $\mu g/m^3$.[32] While these studies are informative and supportive of standard cohort study results, these cohorts and other study populations are somewhat limited either in their geographic scope or the age of the individuals included in the study.

This study examines the use of inverse probability of treatment weights with the Cox proportional hazards model, in which $PM_{2.5}$–mortality associations are estimated with $PM_{2.5}$ measured as a continuous exposure across a large, nationwide sample of US adults. This method primarily accounts for selection bias within the cohort, while also controlling for confounding bias attributable to measured covariates. Under certain statistical assumptions, the estimates provided with inverse probability weighted regression also have a causal interpretation. The widespread use of Cox models in survival analysis, given their ability to stratify baseline hazard estimates, makes them a good candidate to use in causal modeling methods. A variety of model specifications and distributional assumptions are implemented, allowing for further sensitivity analysis of the estimated effects.

Methods

Study population, air pollution data, and data access

The observations used in this study were obtained from the National Health Interview Survey (NHIS), an annual cross-sectional household survey administered by the National Center for Health Statistics (NCHS). This large, nationally representative dataset was constructed of publicly available personal data, with the addition of restricted-use mortality follow-up through 2015 using the National Death Index.[33–35] The cohort includes 635,539 civilian noninstitutionalized individuals aged 18 to 84 and living within the contiguous United States at the time of their interview between 1986 and 2014; these study participants had information available for age, sex, race–ethnicity, educational attainment, marital status, income level, urban–rural designation, census tract, interview date, mortality status, smoking status, body mass index (BMI) information, and date of death (for the deceased). For all-cause mortality, censoring for surviving individuals was set to be the last day of follow-up (31 December 2015), whereas in the cardiopulmonary mortality analysis, deaths to other causes were censored at the date of death. Summary statistics are provided in Table 1. Although the study population is nationally representative, the weighting method as described below generates a pseudo-population in which exposure is disassociated from other measured covariates (which may or may not be confounders); as such, the statistics provided in the table represent the true cohort rather than any pseudo-population used in a weighted analysis.

NCHS employees used restricted-use geographic data to assign estimated long-term pollution exposure values to respondents based on their census tract of residency at the time of interview. Annual pollution exposures were estimated for each census block using national regulatory monitoring data from 1999 to 2015 within a universal kriging model employing land-use regression methods and hundreds of variables.[36] These models include variables such as road density, population density, land use, land cover, and elevation. Cross-validation of the models yielded 10-fold cross-validation R^2 between 0.78 and 0.90. Population-weighted averages of these estimates were estimated at the census tract level to construct a 17-year average (1999–2015) for $PM_{2.5}$ concentration in each census tract to be used as an estimate of long-term exposure. A lack of geographic follow-up data prevented the assignment of pollution from varying as study subjects move post-interview. These modeled air pollution data are publicly available at www.caces.us (the Center for Air, Climate, and Energy Solutions [CACES]), with more detailed descriptions of data estimation and assignment available elsewhere.[17,36] A histogram representing estimated exposure for each individual in this study is presented in Figure 1, along with fitted probability distribution graphs for select distributions.

All analyses were performed at the Research Data Center (RDC) in Hyattsville, MD, with all released results having been previously reviewed and approved to ensure that NHIS survey respondents remain deidentified. The NCHS approved all methods for informed consent, data collection, linkage of the public data to pollution estimates and mortality follow-up, construction of the dataset, and statistical analysis. All information contained in this study originates from deidentified publicly accessible data and is therefore exempt from federal regulations regarding the protection of human research subjects. All findings and conclusions of this study are of the authors alone and are not necessarily representative of the views of the RDC, the NCHS, the Environmental Protection Agency, or the Centers for Disease Control and Prevention.

Table 1.

Cohort summary statistics

Variable	
Total number in cohort	635,539
Total deaths	106,385
Cardiopulmonary	43,195
Gender	
% Male	44.54
% Female	55.46
Age at time of survey (yrs, mean)	45.3
Race/ethnicity	
% Non-Hispanic White	67.51
% Hispanic	14.08
% Non-Hispanic Black	14.01
% All other/unknown	4.40
Income (inflation adjusted to 2015)	
% $0–35,000	38.04
% $35–50,000	15.47
% $50–75,000	18.79
% $75,000+	27.71
Marital status	
% Married	49.57
% Divorced	14.06
% Separated	3.59
% Never married	24.31
% Widowed	8.47
Education	
% <High School grad	18.63
% High School grad	30.37
% Some College	27.10
% College grad	15.03
% >College grad	8.87
Urban/rural	
% Urban	77.64
% Rural	22.36
Census region	
% Northeast	18.08
% Midwest	23.71
% South	35.74
% West	22.46
BMI	
% <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)	10.7, 2.4, 2.5–19.2

BMI, body mass index.

Statistical methods

Inverse probability weighting

The inverse probability weights (IPWs) used in this analysis were generated by taking the inverse of the conditional probability of exposure to a given value in the continuum of $PM_{2.5}$ concentrations and stabilized by multiplying these weights by the marginal probability of the level of exposure. Because this weighted estimation relies heavily on distributional assumptions, several approaches were taken to evaluate the robustness of the results of this analysis. IPWs were generated with multiple distributions: homoscedastic normal, Student's t with 1 and 5 degrees of freedom, and a gamma distribution (which accounts for potential heteroscedasticity through the definition of the mean as a function of its variance), as well as with a quantile binning approach that does not require distributional assumptions, with 10 and 20 distinct bins. Following the analysis by Naimi et al,[37] weights were truncated at the 1st and 99th percentiles of estimated probability of exposure. The parameters of these distributions were estimated from the available data, and conditional distributions used the covariates listed above.

If no unmeasured confounders exist, weighting by IPWs yields a pseudo-population in which exposure is independent from all covariates.[38,39] While the weighted cohorts are no longer representative of the entire adult civilian noninstitutionalized US population, this process allows for estimation of the causal effect of increased $PM_{2.5}$ exposure if there are no unmeasured confounders and other assumptions are satisfied, as further discussed in the supplemental material (S1); http://links.lww.com/EE/A73 this mimics a randomized controlled trial in which all participants are exposed to a continuous treatment rather than a common binary one.[40] This method also adjusts for selection and measured confounding biases, as standard regression adjustment cannot.[41] Unlike covariate adjustment with the propensity score or propensity score matching on discretized variables, regression with IPWs allows for direct computation of meaningful, interpretable estimates.[42] The extent to which these estimates may be viewed as causal relies upon several key assumptions that are discussed in further detail in the supplemental material (S1); http://links.lww.com/EE/A73. A visualization of the relationship of interest may be found in Figure 2, which presents the assumed conceptual relationship between outdoor $PM_{2.5}$ concentrations and mortality; potential confounders of the relationship between both outdoor $PM_{2.5}$ concentrations and personal $PM_{2.5}$ exposure have also been indicated.

Model design

Cox proportional hazards models were used to estimate hazard ratios associated with a 10 $\mu g/m^3$ increase in ambient $PM_{2.5}$ exposure. All models were estimated with the PHREG procedure in SAS (SAS Institute, Cary, North Carolina). Individuals of each age group (18–24, and subsequent 5-year age groups), sex, and race–ethnicity received their own baseline hazard functions, while other covariates were included as confounding variables: income level, educational attainment, marital status, BMI, smoking status, census region, and urban/rural designation (as defined by the US Census Bureau).[43] Each of these covariates, including age group, sex, and race–ethnicity, were included as confounders while constructing the IPWs for weighting the estimated models.

The Cox models used in this analysis are marginal structural models, following the definitions of Robins et al.[38] A weighted model with only $PM_{2.5}$ exposure (1, hereafter denoted as the "IPW model") and a weighted model that also includes the full slate of covariates (2, or "IPW-covariate model") are both estimated.[38,44] The IPW-covariate model is similar to the doubly robust estimator for binary exposure, which is robust to misspecification in either the weight model or the outcome model, but not both.[45]

Variance estimation

Parametric propensity score weighting requires assumptions about the relation between exposure and confounders, such as the distributional form of conditional exposure and the linearity and degree of, and interactions between, confounders. It is common practice to use bootstrapping methods to estimate standard errors and associated confidence interval.[45] In this study, 100 bootstrapped datasets are generated and used to estimate a 95% confidence interval for the true effect estimate. Because the weights are empirically generated, robust variance estimators may also be used—confidence intervals using this method of variance estimation are also provided as a comparison to the bootstrapped confidence interval.[46,47] For reference, confidence intervals generated through the use of naive standard errors are likewise listed.

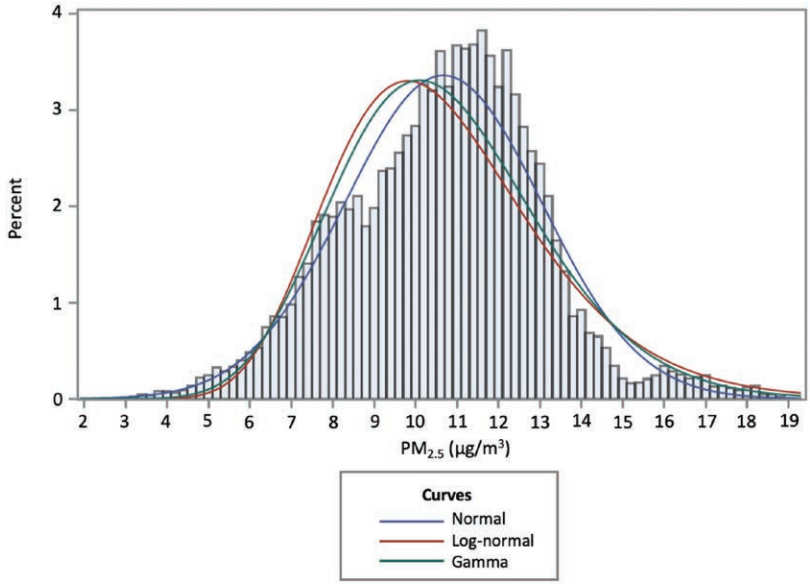

Figure 1. Modeled $PM_{2.5}$ exposure distribution for study population, with select fitted PDFs.

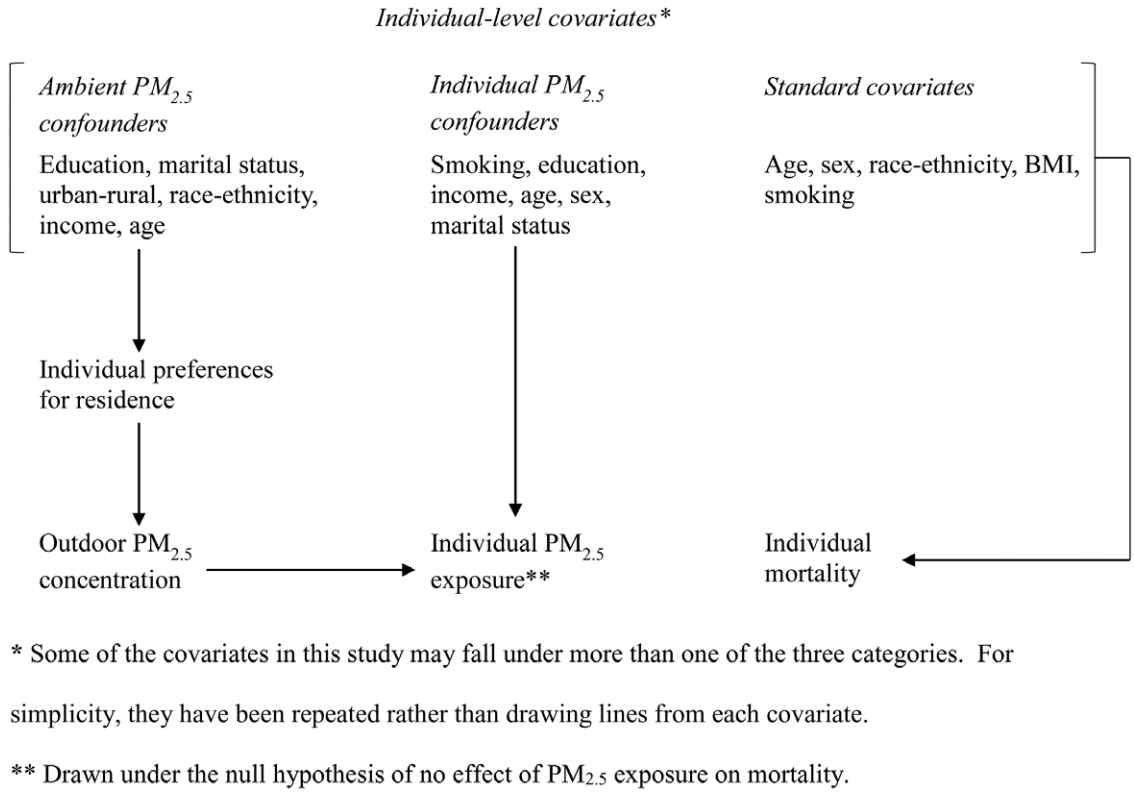

Figure 2. Directional acyclic graph (DAG) of causal pathways affecting individual mortality. *Some of the covariates in this study may fall under more than one of the three categories. For simplicity, they have been repeated rather than drawing lines from each covariate. **Drawn under the null hypothesis of no effect of $PM_{2.5}$ exposure on mortality. BMI, body mass index.

Imbalance in pollution exposure

The pseudo-population generated by the IPW method is designed to be balanced across all measured covariates. Two methods were implemented to assess the need for covariate rebalancing and the degree to which weighting improves this balance across the study population. First, an unweighted linear model was fitted predicting $PM_{2.5}$ exposure with all measured confounders. The R^2 of this linear model is compared with the R^2 from linear models weighted with each of the generated IPWs. The second, and more conventional, method of assessing balance is the testing for equality of standardized covariate means across quantiles of measured exposure.[48,49] This process was adapted for the present analysis as follows. First, the observations were divided into four quartiles of modeled $PM_{2.5}$ exposures. An indicator variable was generated for each possible category

of the previously listed categorical covariates, yielding a total of 33 numerical variables for quantile balance assessment. A t-test was used to test the equality of the means of each indicator variable between two groups—those within a given quartile, and those within the other three quartiles combined. Finally, the number of t-statistics greater than 1.96 (for large degrees of freedom and $\alpha = 0.05$) for each of the 33 variables and each quartile was totaled for each weighting distribution. A reduction in the number of statistically significant standardized differences indicates an improvement in covariate balance.

Results

Covariate balance

As shown in Table 2, $PM_{2.5}$ exposure is correlated with the other covariates included in the model. The R^2 from an unweighted linear regression is 0.1462, which is relatively small but nevertheless indicates a potential confounding effect. The R^2 from each of the weighted linear regressions is smaller than that the unweighted regression; in some cases, such as the Student's t distribution with 5 degrees of freedom ($R^2 = 0.0222$), the reduction is substantial. These reduced values indicate that the stabilized IPWs have the intended effect of improving covariate balance across treatment groups.

The second approach likewise indicates covariate imbalance among the unweighted population, though it does not indicate as significant of an improvement as the first method. The number of t-statistics greater than 1.96 are displayed in Table 2. Without reweighting the population, 116 differences are statistically significant. Using IPWs to test standardized differences, the balance improves only slightly—the number of significant differences ranges from 106 to 114. This apparent lack of improved balance may reflect the discretization of the data, though coupled with the low R^2 values yielded by first approach (even in the unweighted case), it suggests a low degree of variation in exposures for individuals with high factor levels.

Estimates

Hazard ratios (associated with a 10 μg/m^3 increase in $PM_{2.5}$ exposure) and 95% confidence intervals estimated with naive, robust, and bootstrapped standard errors are provided in Tables 3 and 4 for all-cause and cardiopulmonary mortality, respectively. The estimated hazard ratio for all-cause mortality using the unweighted model with only $PM_{2.5}$ included as a variable is 1.178 (robust confidence interval (CI): 1.147, 1.210), while estimates generated by the IPW model with various weighting distributions range from 1.091 to 1.135. For the full model, with all covariates included, the unweighted results yielded a point estimate of 1.126 (robust CI: 1.094, 1.159); the IPW-covariate model estimates range from 1.111 to 1.121. Robust standard errors fall between 0.0143 and 0.0245 for the IPW model and between 0.0151 and 0.0231 for the IPW-covariate model, compared with the corresponding unweighted models' standard errors of 0.0137 and 0.0148, respectively. Bootstrapped and standard hazard ratios are similar for both unweighted and weighted models.

Estimated hazard ratios for cardiopulmonary mortality are 1.329 (robust CI: 1.274, 1.386) for the unweighted model without all covariates, with IPW models producing estimates from 1.214 to 1.260. The unweighted model with all covariates yielded an estimate of 1.242 (robust CI: 1.187, 1.299) compared with IPW-covariate model hazard ratios of 1.227 to 1.235. Robust standard errors exhibited a similar trend as with the all-cause mortality analysis—the unweighted model without covariates included in the regression model yielded a standard error of 0.0215, compared with IPW model standard errors from 0.0225 to 0.0375. For the models with all covariates included, the unweighted model estimated a smaller standard error (0.0231) than the IPW-covariate models (from 0.035 to 0.0352). Although the differences between bootstrapped and standard hazard ratios are slightly larger for cardiopulmonary mortality than for all-cause mortality, these differences are small.

The three methods of variance estimation yielded different standard errors and associated confidence intervals, though each was significant at a 95% confidence level. The naive standard errors are smaller for each of the weighted models than for the unweighted models, although the robust and bootstrapped standard errors are smallest for the unweighted model in each case. The bootstrapped standard errors are generally larger than those generated with the robust variance estimator, while the robust standard errors are always at least weakly larger than the naive standard errors for the corresponding models. A comparison of the standard errors for the log-hazard ratios is provided in Tables 3 and 4 for all-cause and cardiopulmonary mortality.

Summary statistics of the various calculated weights are presented in Table 5. Certain distributions, such as the normal homoscedastic and gamma distributions, yielded more extreme values for the estimated weights. To prevent large biases from potentially misestimated weights, all weights (including those with smaller variance) were truncated at the 1st and 99th percentiles. Histograms of the generated weights are presented in Figure 2, demonstrating the differences between the distributions of the untruncated and truncated weights. As can be seen in Tables 3 and 4, the differences between estimates vary only slightly between the untruncated and truncated weights generated from the same distributions (Figure 3).

Table 2.

Balance assessment of IPWs

Weight type	R^2	Significant standardized differences[a] (out of 132)
Unweighted	0.146	116
Normal	0.059	106
Truncated normal	0.036	113
t, 1 df	0.025	113
Truncated t, 1 df	0.026	113
t, 5 dfs	0.022	114
Truncated t, 5 dfs	0.024	113
Gamma	0.054	109
Truncated gamma	0.033	111
10 quantiles	0.081	112
Truncated 10 quantiles	0.075	110
20 quantiles	0.092	113
Truncated 20 quantiles	0.084	112

[a]Significant difference using a difference-of-means test with a 95% confidence level.
df indicates degree of freedom; t, the Student's t distribution.

Discussion

Marginal structural Cox proportional hazards models used in estimating long-term pollution–mortality associations allow for the analysis of exposure to $PM_{2.5}$ where treatment assignment is disassociated from measured covariates, mimicking a randomized control trial with a weighted pseudo-population. There is evidence of covariate imbalance across quartiles of measured covariates, and all measured covariates are weak, but statistically significant predictors of estimated $PM_{2.5}$ exposure with an unweighted linear regression model ($R^2 = 0.146$). Though the degree of measured correlation between $PM_{2.5}$ exposure and the measured covariates is small, it nonetheless decreases when the study population is weighted by the generated IPWs; in some cases, the R^2 decreases to as little as 0.022 (Table 2). Alternatively, the high dimensionality and discretization of much of the data into binary variables result in significant differences in covariate

Table 3.

All-cause mortality, hazard ratios for 10 μg/m³ increase in $PM_{2.5}$ exposure

			95% Confidence intervals			Standard errors		
Unweighted models	**HRs**	**Bootstrapped HRs**	**Naive**	**Robust**	**Bootstrapped[a]**	**Naive**	**Robust**	**Bootstrapped[a]**
Without control for covariates	1.178	1.180	1.147–1.210	1.147–1.210	1.150–1.212	0.0136	0.0137	0.0134
With control for covariates	1.126	1.126	1.095–1.158	1.094–1.159	1.096–1.158	0.0144	0.0148	0.0140
IPW models (without control for covariates)								
Normal	1.094	1.096	1.069–1.120	1.045–1.145	1.053–1.141	0.0119	0.0232	0.0203
Truncated normal	1.108	1.109	1.080–1.136	1.073–1.144	1.074–1.146	0.0138	0.0164	0.0166
t, 1 df	1.133	1.135	1.103–1.164	1.102–1.165	1.103–1.167	0.0136	0.0143	0.0144
Truncated t, 1 df	1.135	1.136	1.105–1.165	1.103–1.167	1.105–1.169	0.0136	0.0143	0.0144
t, 5 dfs	1.118	1.118	1.090–1.147	1.086–1.152	1.083–1.154	0.0131	0.0151	0.0162
Truncated t, 5 dfs	1.121	1.121	1.093–1.151	1.089–1.154	1.086–1.157	0.0132	0.0149	0.0160
Gamma	1.093	1.093	1.068–1.120	1.042–1.147	1.041–1.148	0.0120	0.0245	0.0250
Truncated gamma	1.104	1.104	1.077–1.133	1.070–1.140	1.067–1.143	0.0129	0.0162	0.0174
10 quantiles	1.092	1.093	1.065–1.119	1.056–1.129	1.058–1.130	0.0126	0.0171	0.0170
Truncated 10 quantiles	1.098	1.099	1.070–1.125	1.063–1.133	1.063–1.136	0.0128	0.0164	0.0170
20 quantiles	1.091	1.092	1.064–1.118	1.054–1.128	1.056–1.130	0.0125	0.0174	0.0172
Truncated 20 quantiles	1.097	1.098	1.070–1.124	1.062–1.133	1.062–1.136	0.0126	0.0166	0.0172
IPW models (with controls for covariates)								
Normal	1.112	1.114	1.086–1.140	1.066–1.161	1.073–1.156	0.0124	0.0219	0.0189
Truncated normal	1.117	1.118	1.089–1.147	1.082–1.154	1.085–1.152	0.0131	0.0165	0.0153
t, 1 df	1.121	1.121	1.091–1.151	1.088–1.154	1.091–1.153	0.0138	0.0151	0.0141
Truncated t, 1 df	1.121	1.122	1.091–1.152	1.089–1.155	1.091–1.153	0.0138	0.0151	0.0142
t, 5 dfs	1.117	1.117	1.088–1.147	1.083–1.152	1.080–1.155	0.0133	0.0156	0.0171
Truncated t, 5 dfs	1.117	1.117	1.088–1.147	1.084–1.152	1.081–1.155	0.0134	0.0154	0.0168
Gamma	1.111	1.111	1.085–1.139	1.062–1.163	1.062–1.162	0.0125	0.0231	0.0229
Truncated gamma	1.112	1.112	1.084–1.142	1.077–1.149	1.074–1.152	0.0132	0.0163	0.0180
10 quantiles	1.117	1.118	1.088–1.146	1.082–1.154	1.085–1.151	0.0132	0.0165	0.0149
Truncated 10 quantiles	1.118	1.118	1.089–1.148	1.083–1.154	1.086–1.152	0.0133	0.0162	0.0150
20 quantiles	1.117	1.117	1.088–1.146	1.081–1.154	1.085–1.151	0.0132	0.0166	0.0151
Truncated 20 quantiles	1.118	1.118	1.089–1.147	1.083–1.154	1.085–1.152	0.0133	0.0162	0.0152

Bootstrapped confidence intervals are generated by using the standard error from 100 bootstrapped samples. Each model controls for age, sex, and race–ethnicity with a nonparametric baseline hazard function.
[a]Confidence intervals and standard errors are for the estimated coefficient or log-hazard ratio.
df, degree of freedom; t, the Student's t distribution.

Table 4.

Cardiopulmonary mortality, hazard ratios for 10 μg/m³ increase in $PM_{2.5}$ exposure

			95% Confidence intervals			Standard errors		
Unweighted models	**HRs**	**Bootstrapped HRs**	**Naive**	**Robust**	**Bootstrapped[a]**	**Naive**	**Robust**	**Bootstrapped[a]**
Without control for covariates	1.329	1.328	1.274–1.387	1.274–1.386	1.274–1.386	0.0215	0.0215	0.0213
With control for covariates	1.242	1.242	1.188–1.299	1.187–1.299	1.184–1.302	0.0227	0.0231	0.0242
IPW models (without control for covariates)								
Normal	1.216	1.215	1.173–1.261	1.134–1.304	1.137–1.297	0.0186	0.0355	0.0336
Truncated normal	1.234	1.231	1.186–1.284	1.174–1.298	1.167–1.298	0.0202	0.0257	0.0271
t, 1 df	1.258	1.256	1.206–1.312	1.204–1.315	1.201–1.315	0.0214	0.0225	0.0231
Truncated t, 1 df	1.260	1.259	1.209–1.315	1.206–1.317	1.203–1.317	0.0215	0.0225	0.0230
t, 5 dfs	1.243	1.241	1.193–1.294	1.186–1.302	1.184–1.301	0.0207	0.0237	0.0242
Truncated t, 5 dfs	1.246	1.245	1.197–1.296	1.191–1.305	1.188–1.305	0.0208	0.0234	0.0239
Gamma	1.214	1.214	1.170–1.260	1.128–1.306	1.135–1.299	0.0188	0.0374	0.0343
Truncated gamma	1.235	1.232	1.187–1.286	1.175–1.299	1.169–1.299	0.0203	0.0255	0.0268
10 quantiles	1.214	1.212	1.168–1.262	1.152–1.280	1.148–1.279	0.0199	0.0268	0.0276
Truncated 10 quantiles	1.221	1.218	1.174–1.270	1.160–1.284	1.156–1.284	0.0201	0.0258	0.0268
20 quantiles	1.214	1.211	1.168–1.261	1.150–1.280	1.146–1.280	0.0196	0.0273	0.0281
Truncated 20 quantiles	1.221	1.219	1.175–1.270	1.160–1.286	1.155–1.286	0.0199	0.0262	0.0273
IPW models (with controls for covariates)								
Normal	1.230	1.229	1.184–1.278	1.152–1.313	1.151–1.312	0.0196	0.0335	0.0334
Truncated normal	1.235	1.232	1.186–1.286	1.174–1.298	1.164–1.302	0.0207	0.0257	0.0286
t, 1 df	1.232	1.230	1.180–1.285	1.176–1.290	1.170–1.293	0.0217	0.0236	0.0253
Truncated t, 1 df	1.232	1.231	1.181–1.286	1.177–1.291	1.171–1.293	0.0217	0.0235	0.0253
t, 5 dfs	1.232	1.230	1.182–1.283	1.174–1.292	1.167–1.295	0.0210	0.0243	0.0266
Truncated t, 5 dfs	1.231	1.229	1.181–1.283	1.174–1.291	1.167–1.294	0.0211	0.0241	0.0264
Gamma	1.227	1.228	1.181–1.276	1.146–1.315	1.149–1.313	0.0197	0.0352	0.0340
Truncated gamma	1.233	1.230	1.184–1.284	1.173–1.296	1.163–1.300	0.0208	0.0255	0.0284
10 quantiles	1.230	1.227	1.181–1.281	1.169–1.294	1.161–1.298	0.0209	0.0258	0.0285
Truncated 10 quantiles	1.229	1.226	1.180–1.281	1.170–1.292	1.161–1.296	0.0210	0.0253	0.0282
20 quantiles	1.230	1.227	1.181–1.281	1.169–1.294	1.160–1.298	0.0207	0.0261	0.0288
Truncated 20 quantiles	1.230	1.227	1.180–1.281	1.170–1.292	1.160–1.297	0.0209	0.0254	0.0285

Bootstrapped confidence intervals are generated by using the standard error from 100 bootstrapped samples. Each model controls for age, sex, and race–ethnicity with a nonparametric baseline hazard function.
[a]Confidence intervals and standard errors are for the estimated coefficient or log-hazard ratio.
df, degree of freedom; t, the Student's t distribution.

Table 5.

Summary statistics for IPWs

Weight type	Mean	Min	Max
Normal	1.012	0.062	114.538
Truncated normal	0.990	0.274	3.090
t, 1 df	1.031	0.318	3.198
Truncated t, 1 df	1.030	0.462	1.964
t, 5 dfs	1.042	0.223	6.509
Truncated t, 5 dfs	1.039	0.383	2.253
Gamma	1.012	0.065	140.586
Truncated gamma	0.988	0.259	2.979
10 quantiles	1.013	0.176	15.512
Truncated 10 quantiles	1.005	0.249	3.179
20 quantiles	1.018	0.132	17.327
Truncated 20 quantiles	1.009	0.236	3.296

df, degree of freedom.

means between exposure groups, regardless of weighting methods. Such covariate imbalance indicates that there may be some degree of bias in the estimated associations between $PM_{2.5}$ and both all-cause and cardiopulmonary mortality risk, which is mitigated by the use of IPWs.

The marginal structural models used in this study supported the original, unweighted estimates of hazard ratios of 1.126 (robust CI: 1.094, 1.159) for all-cause mortality and 1.242 (robust CI: 1.187, 1.299) for cardiopulmonary mortality. All full models weighted by various IPWs yielded point estimates that were smaller in magnitude than the unweighted model, though with universally larger standard errors, as well. IPW models, which controlled only for covariates in the denominator of the stabilized weights, produced lower estimates than the corresponding unweighted models, which only controlled for age groups, sex, and race–ethnicity through a nonparametric baseline hazard function. While IPW models often yielded smaller estimates than the weighted full models, there was no significant

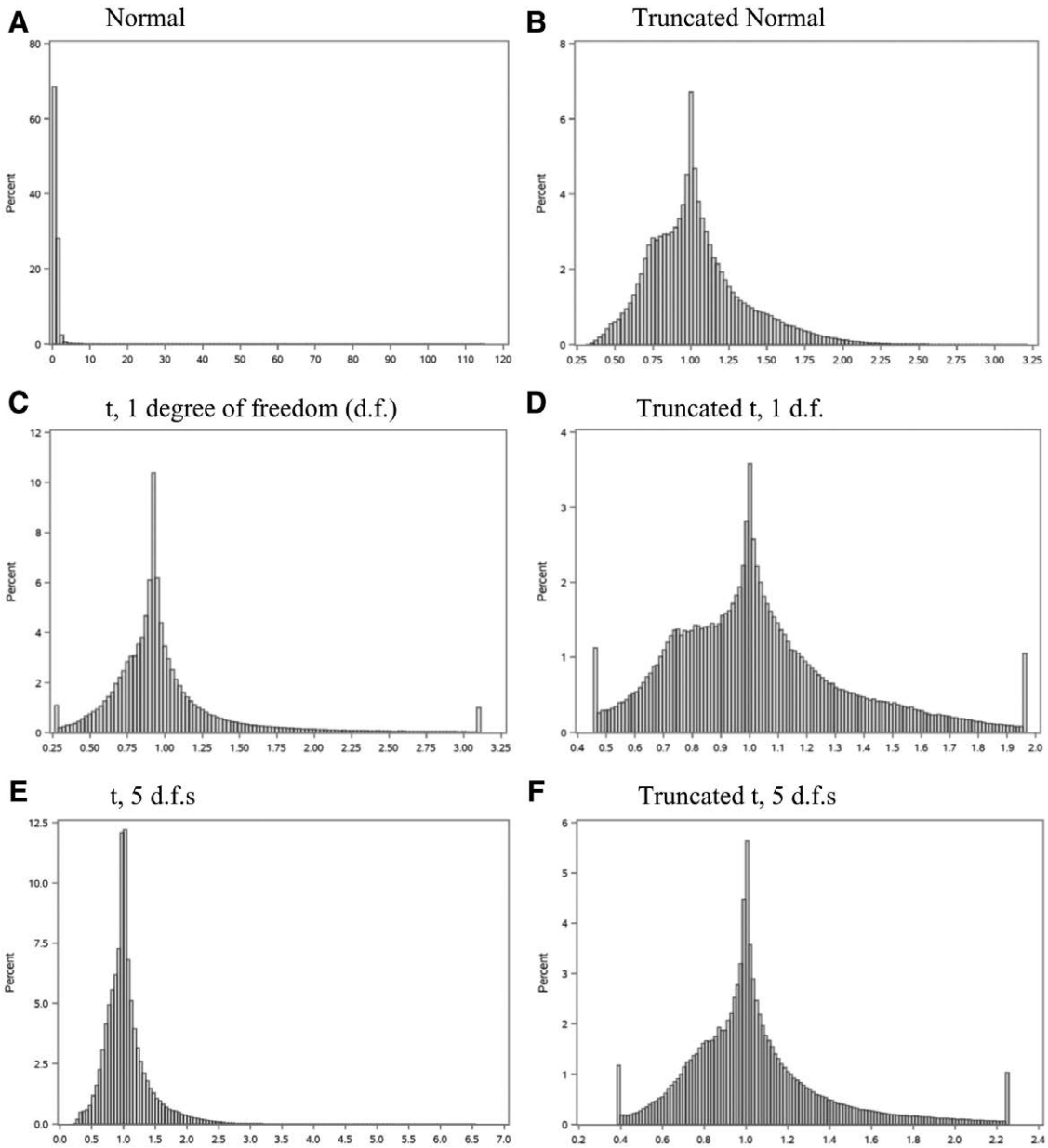

Figure 3. (*Continued*)

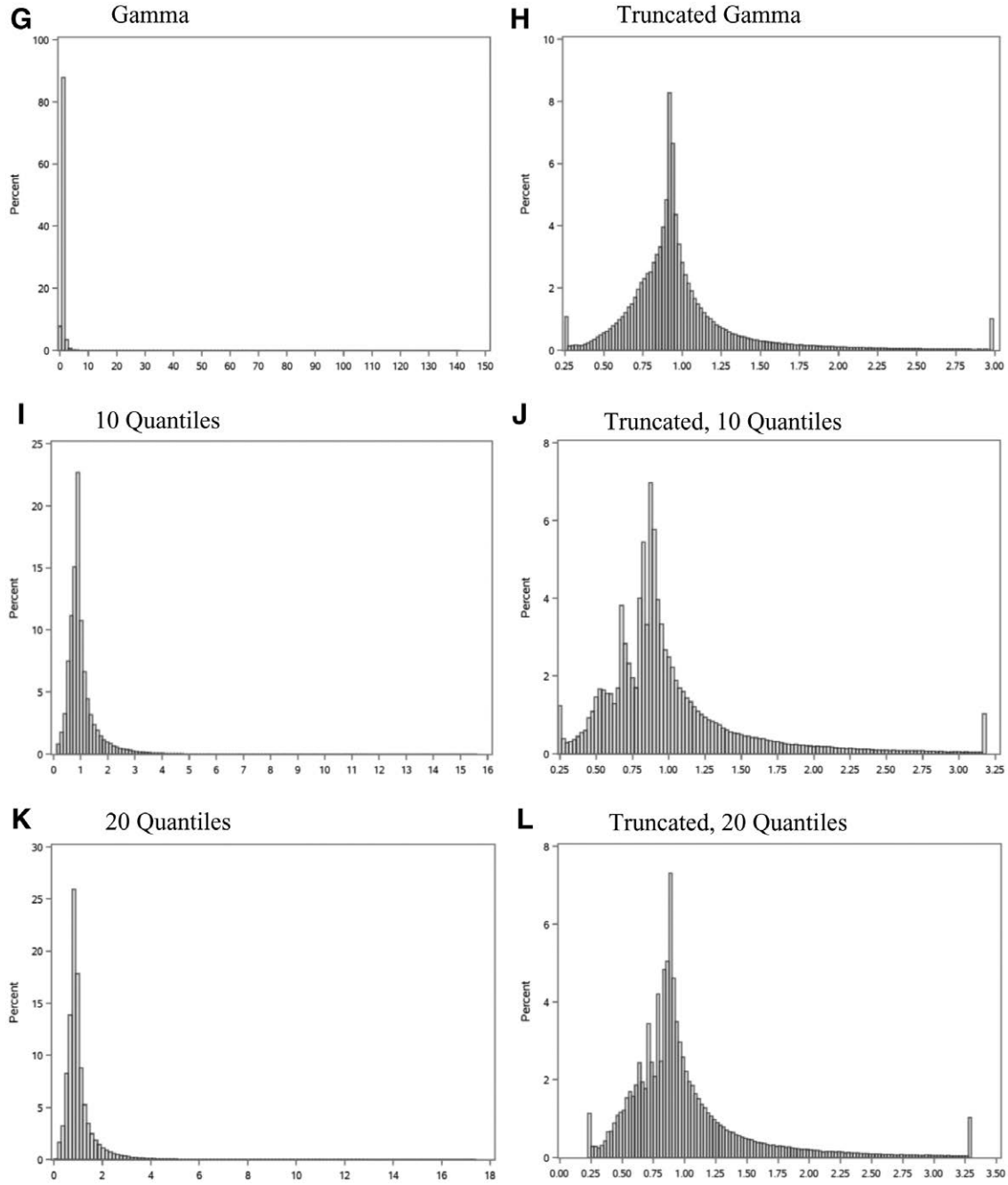

Figure 3. Comparisons of IPW histograms. df, degree of freedom; t, Student's t distribution.

difference between any of these and the unweighted full model; in this setting, the use of IPWs alone provides a reasonable estimate for the $PM_{2.5}$–mortality associations for both all-cause and cardiopulmonary mortality. This similarity is a stark contrast to the hazard ratios estimated by the unweighted models with no covariates included—the estimated hazard ratio for the IPW-covariate model was lower than that of the unweighted model with no covariates by approximately 5% for all-cause and 9% for cardiopulmonary mortality, though the associated confidence intervals overlap for both models' estimates.

The results of this study are comparable to findings by other large cohort studies, such as the hazard ratios for all-cause mortality offered by the Six-cities[1,6] (1.14, CI: 1.07–1.22) and the American Cancer Society Cancer Prevention Study II[2,9,14] (1.07, 1.06–1.09) studies. Significant effects on all-cause mortality are likewise found in the Medicare cohort, though a doubly robust additive hazards model was used rather than a proportional hazards model.[26] The model used in this analysis is similar to the doubly robust additive hazards model by using IPWs and including covariates in the regression, as well; both methods aim to reduce bias in the estimate of the unweighted models. However, the same doubly robust property has not yet been proven for Cox models following this form.

Though the use of inverse probability weighting requires the correct specification of the conditional exposure distribution, the wide array of both untruncated and truncated weights generated from different distributions and quantile binning approaches suggests that point estimates of the hazard ratios are relatively insensitive to the choice of distribution. Even for weights which take on extreme values, such as the normal and

gamma distributions, there is little variation in the point estimates between the truncated and untruncated weights. However, the different IPWs, and truncated weights with extreme values, do result in markedly different confidence intervals for some models; for example, the confidence interval for the IPW-covariate model estimates for all-cause mortality using the normal weights (1.112, robust CI: 1.066, 1.161) is larger than that of the model estimated with truncated normal weights (1.117, robust CI: 1.082, 1.154). Bootstrapped standard errors and confidence intervals display a similar pattern. Although there is some degree of variation in both estimated hazard ratios and standard errors, with their associated confidence intervals, all estimated associations are significant at a 95% confidence level. This suggests that after controlling for confounders within the model itself, there is little residual treatment assignment bias; the similarity in estimates—whether confounders are accounted for in the weights, model, or both—mirrors the properties of the doubly robust model.[26]

This analysis does not account for copollutants such as ozone, that have been included in similar pollution-related mortality studies. Several studies have examined models with two and three pollutants, consistently reporting an association between $PM_{2.5}$ and early mortality even when controlling for other airborne pollutants. A recent analysis by Lefler et al[52] explored one-, two-, and six-pollutant models of early mortality and pollution exposure using the same NHIS dataset used in the present analysis. $PM_{2.5}$ was consistently associated with early mortality even after including $PM_{2.5–10}$, SO_2, NO_2, O_3, and CO both pairwise and all together (respectively, particulate matter from 2.5 to 10 μm in aerodynamic diameter, sulfur dioxide, nitrogen dioxide, ozone, and carbon monoxide). Although SO_2 and $PM_{2.5–10}$ concentrations were associated with mortality risk in the NHIS data, the associations were smaller and less robust than the association with $PM_{2.5}$. The relationship between $PM_{2.5}$ exposure and mortality risk was not highly sensitive to controlling for SO_2 and $PM_{2.5–10}$ in multipollutant models.

While the approach used in the present analysis accounts for confounding by measured covariates, it fails to adjust for potential bias due to omitted or insufficiently controlled for factors that may be associated with both mortality risk and measured $PM_{2.5}$ exposure. Although the measured covariates included in the model span a wide variety of potential confounders, it is possible that there remain some unknown and unmeasured confounders. The increased hazard ratios in the weighted models when moving from an IPW model to an IPW-covariate model suggest that the addition of further covariates would have a minimal effect, as the most important covariates have been included in the models. Furthermore, stepwise sensitivity analyses using unweighted Cox models on this data indicated that results for unweighted models were not sensitive to the choice of covariates included in the model.[17] Another limitation is that direct measures of long-term exposure to $PM_{2.5}$ are not used in this study—$PM_{2.5}$ was only monitored throughout the entire United States beginning in 1999, meaning that those who were surveyed before may have been exposed to more pollution at the time of their survey. Furthermore, each individual's location at the time of their survey was assumed to be their residence over the course of the study, as no geographic follow-up or indication of relocation was provided in the NHIS data. The lack of follow-up for other covariates in the NHIS data was also a limitation of this study, as it prevented for controls of time-varying information for variables such as income. This analysis also assumes that the spatial variation of $PM_{2.5}$ concentrations has been constant over time. Additionally, with the exception of geographic and temporal terms in the models, only individual-level risk factors were included in these models; this weakens the assumption of no unmeasured confounders, although several individual-level variables such as income and education act as a proxy for confounders that have a causal impact on $PM_{2.5}$ concentrations. The extent to which these estimates may be viewed as causal is also dependent on the extent to which key assumptions in causal inference are satisfied; more details about these assumptions and support for their plausibility may be found in the supplemental material (S1); http://links.lww.com/EE/A73.

This study furthers the use of propensity score and causal modeling methods in examining associations between long-term $PM_{2.5}$ exposure and mortality. The use of a large, nationally representative dataset allows for both control and covariate balance assessment on a number of variables, including smoking status and BMI data. Multiple distributions and weight generation techniques, such as quantile binning, were used in this study to account for several distributional assumptions, nonparametric estimation of propensity scores, potential heteroscedasticity, and possible thicker tails in the exposure distributions. The results demonstrate the robustness of the unweighted model and relative insensitivity to the choice of IPW that is used in each model. These findings contribute to a growing body of evidence suggesting that the estimated $PM_{2.5}$–mortality associations are causal in nature; given the prevalence of ambient $PM_{2.5}$ air pollution, these results have significant implications for general public health.

ACKNOWLEDGMENTS

This publication was developed as part of the Center for Clean Air Climate Solutions (CACES), which was supported under Assistance Agreement No. R835873 awarded by the US Environmental Protection Agency. It has not been formally reviewed by Environmental Protection Agency (EPA). The views expressed in this document are solely those of authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication.

Conflicts of interest statement

The authors declare that they have no conflicts of interest with regard to the content of this report.

REFERENCES

1. Dockery DW, Pope CA 3rd, Xu X, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med. 1993;329:1753–1759.
2. Pope CA 3rd, Burnett RT, Thun MJ, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA. 2002;287:1132–1141.
3. Miller KA, Siscovick DS, Sheppard L, et al. Long-term exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med. 2007;356:447–458.
4. Puett RC, Hart JE, Suh H, Mittleman M, Laden F. Particulate matter exposures, mortality, and cardiovascular disease in the health professionals follow-up study. Environ Health Perspect. 2011;119:1130–1135.
5. Lipsett MJ, Ostro BD, Reynolds P, et al. Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort. Am J Respir Crit Care Med. 2011;184:828–835.
6. 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 1974 to 2009. Environ Health Perspect. 2012;120:965–970.
7. Crouse DL, Peters PA, Hystad P, et al. Ambient PM2.5, 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. 2015;123:1180–1186.
8. Hart JE, Liao X, Hong B, et al. The association of long-term exposure to PM2.5 on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. Environ Health. 2015;14:38.
9. Pope CA 3rd, Turner MC, Burnett RT, et al. Relationships between fine particulate air pollution, cardiometabolic disorders, and cardiovascular mortality. Circ Res. 2015;116:108–115.
10. Villeneuve PJ, Weichenthal SA, Crouse D, et al. Long-term exposure to fine particulate matter air pollution and mortality among Canadian Women. Epidemiology. 2015;26:536–545.

11. Pinault L, Tjepkema 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.
12. Pinault LL, Weichenthal S, Crouse DL, et al. Associations between fine particulate matter and mortality in the 2001 Canadian Census Health and Environment Cohort. Environ Res. 2017;159:406–415.
13. Thurston GD, Ahn J, Cromar KR, et al. Ambient particulate matter air pollution exposure and mortality in the NIH-AARP diet and health cohort. Environ Health Perspect. 2016;124:484–490.
14. Jerrett M, Turner MC, Beckerman BS, et al. Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates. Environ Health Perspect. 2017;125:552–559.
15. Di Q, Wang Y, Zanobetti A, et al. Air pollution and mortality in the medicare population. N Engl J Med. 2017;376:2513–2522.
16. Parker JD, Kravets N, Vaidyanathan A. 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. Circ. 2018;137:1688–1697.
17. Pope CA, Lefler JS, Ezzati M, et al. Mortality risk and fine particulate air pollution in a large, representative cohort of U.S. Adults. Environ Health Perspect. 2019; 127:1–9.
18. Carey IM, Atkinson RW, Kent AJ, van Staa T, Cook DG, Anderson HR. Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. Am J Respir Crit Care Med. 2013;187:1226–1233.
19. Cesaroni G, Badaloni C, Gariazzo C, et al. Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. Environ Health Perspect. 2013;121:324–331.
20. Beelen R, Raaschou-Nielsen O, Stafoggia M, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. Lancet. 2014;383:785–795.
21. Fischer PH, Marra M, Ameling CB, et al. Air pollution and mortality in seven million adults: the Dutch Environmental Longitudinal Study (DUELS). Environ Health Perspect. 2015;123:697–704.
22. Bentayeb M, Wagner V, Stempfelet M, et al. Association between long-term exposure to air pollution and mortality in France: a 25-year follow-up study. Environ Int. 2015;85:5–14.
23. Tseng E, Ho WC, Lin MH, Cheng TJ, Chen PC, Lin HH. Chronic exposure to particulate matter and risk of cardiovascular mortality: cohort study from Taiwan. BMC Public Health. 2015;15:936.
24. Yin P, Brauer M, Cohen A, et al. Long-term fine particulate matter exposure and nonaccidental and cause-specific mortality in a large national cohort of Chinese Men. Environ Health Perspect. 2017;125:117002.
25. Ebenstein A, Fan M, Greenstone M, He G, Zhou M. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. Proc Natl Acad Sci U S A. 2017;114:10384–10389.
26. Wang Y, Lee M, Liu P, et al. Doubly robust additive hazards models to estimate effects of a continuous exposure on survival. Epidemiology. 2017;28:771–779.
27. Wu X, Braun D, Kioumourtzoglou MA, Choirat C, Di Q, Dominici F. Causal inference in the context of an error prone exposure: air pollution and mortality. Ann Appl Stat. 2019;13:520–547.
28. Wang Y, Kloog I, Coull BA, Kosheleva A, Zanobetti A, Schwartz JD. Estimating causal effects of long-term PM2.5 exposure on mortality in New Jersey. Environ Health Perspect. 2016;124:1182–1188.
29. Kioumourtzoglou MA, Schwartz J, James P, Dominici F, Zanobetti A. PM2.5 and mortality in 207 US cities: modification by temperature and city characteristics. Epidemiology. 2016;27:221–227.
30. Schwartz J, Fong K, Zanobetti A. A national multi-city analysis of the causal effect of local pollution, NO2, and PM2.5 on mortality. Environ Health Perpect. 2018;126:087004.
31. Schwartz JD, Wang Y, Kloog I, Yitshak-Sade M, Dominici F, Zanobetti A. Estimating the effects of PM2.5 on life expectancy using causal modeling methods. Environ Health Perspect. 2018;126:127002.
32. Makar M, Antonelli J, Di Q, Cutler D, Schwartz J, Dominici F. Estimating the causal effect of low levels of fine particulate matter on hospitalization. Epidemiology. 2017;28:627–634.
33. National Center for Health Statistics (NCHS). 2015. 2014 National Health Interview Survey: Survey Description. Hyattsville, Maryland: NHIS, Division of Health Interview Statistics. Available at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/ Dataset_Documentation/NHIS/2014/srvydesc.pdf. Accessed 31 August 2018.
34. – NCHS. 2018a. National Health Interview Survey, 1986–2014. NHIS data, questionnaires and related documentation. Available at: https://www.cdc.gov/nchs/nhis/data-questionnaires- documentation.htm. Accessed 31 August 2018.
35. NCHS. 2018b. NCHS data linked to NDI mortality files. https://www.cdc.gov/nchs/ data-linkage/mortality.htm. Accessed 31 August 2018.
36. Kim S-Y, Bechle M, Hankey S, Sheppard EA, Szpiro AA, Marshall JD. 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.
37. Naimi AI, Moodie EE, Auger N, Kaufman JS. Constructing inverse probability weights for continuous exposures: a comparison of methods. Epidemiology. 2014;25:292–299.
38. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11:550–560.
39. Coffman DL, Zhong W. Assessing mediation using marginal structural models in the presence of confounding and moderation. Psychol Methods. 2012;17:642–664.
40. Stürmer T, Rothman KJ, Glynn RJ. Insights into different results from different causal contrasts in the presence of effect-measure modification. Pharmacoepidemiol & Drug Saf. 2006;15:698-709.
41. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008;168:656–664.
42. Austin PC. An Introduction to propensity score methods for reducing the effects of confounding in observational Studies. Multivariate Behav Res. 2011;46:399–424.
43. U.S. Census Bureau. 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=7a41374f6b03456e9d-138cb014711e01. Accessed 16 April 2019.
44. Karim ME, Gustafson P, Petkau J, et al. Marginal structural Cox models for estimating the association between β-interferon exposure and disease progression in a multiple sclerosis cohort. Am J Epidemiol. 2014;180:160–171.
45. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. Am J Epidemiol. 2011;173:761–767.
46. Buchanan AL, Hudgens MG, Cole SR, Lau B, Adimora AA; Women's Interagency HIV Study. Worth the weight: using inverse probability weighted Cox models in AIDS research. AIDS Res Hum Retroviruses. 2014;30:1170–1177.
47. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. J Am Stat Assoc. 1989;84:1074-1078.
48. Hirano, K, Imbens GW. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Serv Outcomes Res Methodol. 2001;2:259–278.
49. Hirano K, Imbens GW. The propensity score with continuous treatments. In: Gelman A, Meng XL eds. Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives. Wiley; 2004:73–84.