Original Research Article

ISEE Environmental Epidemiology

OPEN

Epidemiological studies likely need to consider $PM_{2.5}$ composition even if total outdoor $PM_{2.5}$ mass concentration is the exposure of interest

Scott Weichenthal[a,*], Tanya Christidis[b], Toyib Olaniyan[b], Aaron van Donkelaar[c], Randall Martin[c], Michael Tjepkema[b], Rick T. Burnett[d], Michael Brauer[d,e]

Background: Outdoor fine particulate air pollution, <2.5 μm ($PM_{2.5}$) mass concentrations can be constructed through many different combinations of chemical components that have varying levels of toxicity. This poses a challenge for studies interested in estimating the health effects of total outdoor $PM_{2.5}$ (i.e., how much $PM_{2.5}$ mass is present in the air regardless of composition) because we must consider possible confounders of the version of treatment-outcome relationships.

Methods: We evaluated the extent of possible bias in mortality hazard ratios for total outdoor $PM_{2.5}$ by examining models with and without adjustment for sulfate and nitrate in $PM_{2.5}$ as examples of potential confounders of version of treatment-outcome relationships. Our study included approximately 3 million Canadians and Cox proportional hazard models were used to estimate hazard ratios for total outdoor $PM_{2.5}$ adjusting for sulfate and/or nitrate and other relevant covariates.

Results: Hazard ratios for total outdoor $PM_{2.5}$ and nonaccidental, cardiovascular, and respiratory mortality were overestimated due to the confounding version of treatment-outcome relationships, and associations for lung cancer mortality were underestimated. Sulfate was most strongly associated with nonaccidental, cardiovascular, and respiratory mortality suggesting that regulations targeting this specific component of outdoor $PM_{2.5}$ may have greater health benefits than interventions targeting total $PM_{2.5}$.

Conclusions: Studies interested in estimating the health impacts of total outdoor $PM_{2.5}$ (i.e., how much $PM_{2.5}$ mass is present in the air) need to consider potential confounders of the version of treatment-outcome relationships. Otherwise, health risk estimates for total $PM_{2.5}$ will reflect some unknown combination of how much $PM_{2.5}$ mass is present in the air and the kind of $PM_{2.5}$ mass that is present.

Introduction

Outdoor fine particle mass concentrations (fine particulate air pollution, <2.5 μm [$PM_{2.5}$]) are regulated around the world owing to decades of epidemiological data supporting important impacts on human health.[1,2] While the term $PM_{2.5}$ suggests a single entity, this metric captures all airborne particles in the $PM_{2.5}$ size range and does not consider chemical composition. As a result, the primary exposure of interest in epidemiological studies designed to support regulatory interventions is total undifferentiated outdoor $PM_{2.5}$ mass concentration (i.e., how much $PM_{2.5}$ mass is present in the air without concern for composition). Expressed in terms of the consistency assumption under the potential outcomes framework,[3] these studies implicitly assume that:

$Y_j^{obs} = Y_j(x, k)$ if $x = X_j$, no matter the value of k.

This relationship states that we assume that the potential outcome observed for individual *j* (i.e., Y_j^{obs}) is a function of the exposure, $x = X_j$, administered by method k, where all k

What this study adds?

Outdoor fine particulate air pollution, <2.5 μm ($PM_{2.5}$) has an important impact on human health, but little attention has focused on the consequences of spatial/temporal differences in $PM_{2.5}$ composition in estimating the health impacts of total outdoor $PM_{2.5}$ mass concentrations. This is an important question, as current regulatory metrics are concerned only with how much $PM_{2.5}$ mass is present in the air and not what kind of $PM_{2.5}$ is present. We examined potential bias in health risk estimates for total outdoor $PM_{2.5}$ (i.e., how much $PM_{2.5}$ is present) after adjusting for potential confounders of the version of treatment-outcome relationships using sulfate and nitrate as two specific examples. Residual analysis was also conducted (i.e., removing variations in total outdoor $PM_{2.5}$ caused by sulfate/nitrate). Our findings suggest that studies interested in the health effects of total outdoor $PM_{2.5}$ need to consider the kind of $PM_{2.5}$ and potential confounders of the version of treatment-outcome relationships. Otherwise, health risk estimates for total outdoor $PM_{2.5}$ will reflect some unknown combination of how much $PM_{2.5}$ mass is present and the kind of $PM_{2.5}$ mass that is present.

[a]Department of Epidemiology, McGill University; Montreal, Canada; [b]Statistics Canada; Ottawa, Canada; [c]Washington University in St Louis; Saint Louis; [d]Institute for Health Metrics and Evaluation; University of Washington, Seattle; and [e]University of British Columbia; Vancouver, Canada

The research described in this article was conducted under contract with the Health Effects Institute (HEI), an organization jointly funded by the United States Environmental Protection Agency (EPA) and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers.

The CanCHEC cohort data can be accessed through Research Data Centers located across Canada conditional on completion of the procedures put in place by Statistics Canada.

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

Corresponding Author. Address: Department of Epidemiology, McGill University, School of Population and Global Health, Biostatistics, and Occupational Health, Montreal, QC, Canada. E-mail: scott.weichenthal@mcgill.ca (S. Weichenthal).

Environmental Epidemiology (2024) 8:e317

Received 11 January, 2024; Accepted 24 May, 2024

Published online 16 July 2024

DOI: 10.1097/EE9.0000000000000317

methods of administering exposure are equivalent. For outdoor $PM_{2.5}$ mass concentrations, this means that we are assuming that particle composition does not matter and that all versions (k) of a given outdoor $PM_{2.5}$ mass concentration are equally harmful. Put another way, we are assuming that two interventions that reduce outdoor $PM_{2.5}$ mass concentrations by 1 $\mu g/m^3$ would have the same health benefits even if one intervention achieved this reduction by reducing sea salt in $PM_{2.5}$ and the other reduced fossil fuel/coal burning-related $PM_{2.5}$, which have been shown to be strongly associated with mortality.[4–6] Intuitively, it seems likely that epidemiological studies of outdoor $PM_{2.5}$ mass concentrations will violate the consistency assumption because some components are known to be more harmful than others. In essence, when we conduct cohort studies of total outdoor $PM_{2,5}$ mass concentrations and mortality, we are contrasting the survival experience of people who are exposed to the kind of $PM_{2.5}$ that is present at higher mass concentrations to the kind of $PM_{2.5}$ that is present at lower mass concentrations without ever specifying what "kind" means. The health risks estimated in these studies (e.g., mortality hazard ratios [HRs]) are generally attributed entirely to differences in outdoor $PM_{2.5}$ mass concentrations (i.e., the "how much" effect) but in reality, differences in health risks across gradients of total outdoor $PM_{2.5}$ mass concentrations reflect some unknown combination of differences both in how much $PM_{2.5}$ is present in the air as well as the kind of $PM_{2.5}$ that is present.

This issue is related to the so-called "no-multiple-versions-of-treatment" assumption in the potential outcomes framework for causal inference,[7] and can result in bias in health risk estimates for outdoor $PM_{2,5}$ if there are unmeasured confounders of "version of treatment"-outcome relationships.[8] While several studies have applied causal inference methods to describe and address the issue of confounding bias in air pollution epidemiology,[9–11] much less attention has focused on the implications of violations of the consistency assumption with respect to estimating the health impacts of total outdoor $PM_{2.5}$ (i.e., how much $PM_{2.5}$ mass is present without concern for composition—the current regulatory metric). This is an important question given that $PM_{2.5}$ composition and toxicity likely vary spatially, temporally, and across the $PM_{2.5}$ distribution. We recently explored the possibility of confounding bias caused by violations of the "no-multiple-versions-of-treatment assumption" in cohort studies of outdoor $PM_{2.5}$ mass concentrations and mortality using simulations.[8] Here we extend this work to a cohort study using the Canadian Census Health and Environment Cohort (CanCHEC) including estimates of total outdoor $PM_{2.5}$ mass concentrations as well as two specific chemical components estimated in outdoor $PM_{2.5}$ (sulfate and nitrate) on a national scale.[12] These components were selected because: (1) They reflect specific chemical compounds in $PM_{2.5}$ (i.e., versions of treatment); (2) Model predictions for these components are strongly correlated with ground-level measurements on a national scale;[12] and (3) These components have been independently associated with mortality in previous studies across the United States.[13] Unlike many previous studies, in this analysis, our goal was not to identify the specific $PM_{2.5}$ components that are most strongly associated with health outcomes. Recent studies have examined this question in detail and confirm that heterogeneity exists in the strengths of associations between specific components in outdoor $PM_{2.5}$ and mortality, including sulfate and nitrate.[13] Instead, our aim was to estimate the magnitude and direction of potential bias in mortality HRs for total outdoor $PM_{2,5}$ mass concentrations (i.e., how much $PM_{2.5}$ mass is present in the air, the regulatory metric of interest) by adjusting for possible confounders of "version-of-treatment"-outcome relationships using the two specific $PM_{2.5}$ components noted above as indicators of "version of treatment." This question is of direct regulatory significance as policy interventions addressing the current regulatory definition of outdoor $PM_{2.5}$ require estimates of the health impacts of total outdoor $PM_{2.5}$ (i.e., how much $PM_{2.5}$ mass is present) that are not entangled with health impacts of the kind of $PM_{2.5}$ mass that is present.

Methods

Study population

Our study population was based on the 2006 CanCHEC. This cohort is made up of respondents to the 2006 long-form census questionnaire, which captures individual and household sociodemographic data that was subsequently linked to longitudinal vital statistics and tax records (which are used for geocoding residential location). Noninstitutionalized respondents to the long-form questionnaire who lived in Canada were considered in scope for linkage.[14] Record linkage and data sharing at Statistics Canada are governed by the Directive on Microdata Linkage and were performed through the Social Data Linkage Environment through deterministic or probabilistic linkages.[15] The 2006 census asked respondents to consent to record linkage with tax files, which allows researchers to attach environmental covariates to a dynamic postal code history. Of the 5,871,337 persons included in the 2006 CanCHEC, 3,776,376 consented to linkage. Consent was less likely in people who were living in the territories of Nunavut, Northwest Territories, and Yukon or reporting Inuit or Indigenous identity. We further excluded persons who were younger than 25 or older than 89 or immigrated to Canada less than 10 years before the census year, which left 3,027,399 people.

The CanCHEC dataset was further customized to create an analytical file for this work, following methods used in past studies.[16,17] This process involved: (1) imputing postal codes, (2) removing business postal codes, (3) recoding census variables, (4) attaching air pollutant concentrations, (5) calculating 10-year moving averages with a 1-year lag for all air pollution estimates, (6) attaching ecological covariates, (7) transforming the file from a person-file to a person-year file, and (8) excluding person-years that did not meet study criteria. Specifically, we performed a postal code imputation when there were gaps in postal code history preceded and proceeded by postal codes that had shared characters.[18] For the 3,027,399 persons included in the preliminary dataset, 92.54% of the person-years (from years 2006–2019) had a complete postal code, 6.56% had no postal code, and 0.90% had a partial (imputed) postal code. Person-years were then excluded if they did not have a full or partial postal code. We also removed postal codes that were associated with a business rather than a residence.[19] Additional person-years were excluded if age during follow-up exceeded 89 years; if they occurred after a person's death; or if postal codes could not be matched to an air pollution estimate, a Canadian Marginalization Index value, or airshed. Person-years with $PM_{2.5}$ values in the 0.005 and 99.5th centile or invalid air pollution estimates were deleted. After performing the exclusions listed above, 3,007,441 unique persons and 35,436,475 person-years were available for analysis.

Outdoor air pollution concentrations

Outdoor air pollution concentrations were assigned as 10-year moving averages with a 1-year lag based on previous analysis examining different time windows of $PM_{2.5}$ exposure.[20] Outdoor $PM_{2.5}$ mass concentrations were assigned to years 1996 and 2019, and estimates for $PM_{2.5}$ components (i.e., sulfate and nitrate) were assigned to years 2000–2016 as these were the years of data available (2016 values were assigned to years 2017–2019). At the start of the follow-up in 2006, the 10-year moving average was informed by 10 years of $PM_{2.5}$ data and 5 years of component data, with subsequent years having additional datapoints to inform the average. $PM_{2.5}$ mass concentrations and $PM_{2.5}$ component fraction data reflect data

versions V4.NA.02.MAPLE and V4.NA.03 available from Washington University in St. Louis (https://sites.wustl.edu/acag/datasets/surface-pm2-5/). These datasets combine multiple satellite retrievals, chemical transport model output, and ground-based observations to predict component and total $PM_{2.5}$ concentrations over North America at a monthly timescale, gridded at approximately 1 km^2 resolution.[12] Component data were also available for ammonium in $PM_{2.5}$ but the ammonium content of outdoor $PM_{2.5}$ was highly correlated with both sulfate (r = 0.89) and nitrate (r = 0.87) and thus was excluded from the analysis.

Statistical analysis

Cox proportional hazard models[21] were used to estimate HRs for total outdoor $PM_{2.5}$ mass concentrations and mortality, adjusting for covariates described below. Individuals were followed from the census date of 2006 until either reaching the age of 89, the year of death, or the end of follow-up in 2019. Mortality outcomes considered were nonaccidental mortality (international classification of diseases-10 [ICD-10] codes A–R), cardiovascular (ICD-10 codes I10–I69), nonmalignant respiratory disease (ICD-10 codes J00–J99), and lung cancer (ICD-10 codes C33–C34). Multiple mortality outcomes were examined as it is possible that the magnitude and direction of potential confounding bias varies by outcome depending on the magnitude/direction of association between sulfate/nitrate in $PM_{2.5}$ and a given outcome.

Three main models were examined: (1) models including total outdoor $PM_{2.5}$ mass concentrations alone; (2) models including total outdoor $PM_{2.5}$ and sulfate or nitrate separately (as potential confounders of the version of treatment-outcome relationships); and (3) models including total outdoor $PM_{2.5}$ with both components together. As sensitivity analyses, additional models were examined, including $PM_{2.5}$, both components, and O_x (redox weighed average of NO_2 and O_3), which has been associated with mortality in previous studies.[17] In addition, models were examined replacing total outdoor $PM_{2.5}$ mass concentrations with the residuals of linear models regressing outdoor $PM_{2.5}$ concentrations on concentrations of sulfate, nitrate, or both (i.e., to capture variation in outdoor $PM_{2.5}$ concentrations not explained by sulfate or nitrate alone or together). All models were stratified by age (5-year age groups), sex, and immigrant status. Models were adjusted for individual-level covariates derived from census respondents (income quintile, visible minority status, Indigenous identity, educational attainment, labor force status, marital status, and occupation) and ecological covariates assigned using census geographies or postal codes (community size, airshed, urban form, material deprivation, residential instability, dependency, and ethnic concentration).[22–24] The directed acyclical graph shown in Supplemental Figure 1; http://links.lww.com/EE/A286 depicts the assumed structural relationships between variables included in our models (using the sulfate example). All HRs were scaled by the interquartile range of each pollutant: $PM_{2.5}$ (3.54 μg/m^3), sulfate (1.13 μg/m^3), nitrate (0.675 μg/m^3).

To characterize the shape of associations between outdoor $PM_{2.5}$ components, total outdoor $PM_{2.5}$ mass concentrations, and mortality, we fit restricted cubic splines (RCS) defined by the number of knots in SAS EG (SAS Institute Inc, Cary, NC). We tested 3–15 knots and selected the minimum AIC from these results. Using the RCS parameter estimates and the covariance matrix from this best-fitting model, we simulated realizations of the RCS at all concentrations between the minimum to the maximum by increments of 0.1 μg/m^3. We estimated the shape of the relationship between the components of interest (sulfate or nitrate) and mortality outcomes while including total outdoor $PM_{2.5}$ mass concentrations in the models (along with the other covariates described above).

A note on interpretation

In this analysis, our goal was to estimate the average effect of long-term exposure to total outdoor $PM_{2.5}$ mass concentrations on mortality (using HRs) taking into consideration the fact that all outdoor $PM_{2.5}$ mass does not have the same composition and toxicity. If there is no bias caused by confounders of the version of the treatment-outcome relationship (and confounders of the treatment-outcome relationship are also controlled for), differences in mortality rates across levels of total outdoor $PM_{2.5}$ mass concentrations (i.e., how much mass is present in the air) are interpreted as reflecting randomly assigned versions of treatment from the distributions of versions of treatment available for total outdoor $PM_{2.5}$ occurring in the study population.[25] In this example, because we are including $PM_{2.5}$ sulfate and nitrate as possible confounders of the version of treatment-outcome relationships, HRs for total outdoor $PM_{2.5}$ mass concentrations reflect health risks at fixed levels of sulfate and nitrate. This interpretation is more consistent with the current regulatory definition of outdoor $PM_{2.5}$ which is concerned only with how much mass is present (and not composition), even if this is not always explicitly stated. If we do not adequately control for confounders of the version of treatment-outcome relationships, differences in mortality rates across levels of total outdoor $PM_{2.5}$ mass concentrations reflect some combination of how much $PM_{2.5}$ mass is present in the air and what kind of $PM_{2.5}$ mass is present across the gradient.

Results

Descriptive data for cohort characteristics are shown in Supplemental Table 1; http://links.lww.com/EE/A286 and outdoor air pollution concentrations are shown in Table 1. In total, 324,548 nonaccidental, 89,700 cardiovascular, 30,009 respiratory, and 34,100 lung cancer deaths were observed during follow-up. Long-term average total outdoor $PM_{2.5}$ concentrations ranged from 2.49 to 14.47 μg/m^3 whereas sulfate concentrations ranged from 0.19 to 5.32 μg/m^3 and nitrate concentrations ranged from 0.0 to 3.45 μg/m^3. Mass proportions of sulfate and nitrate in total outdoor $PM_{2.5}$ are shown in Figure 1 and varied both within levels of total outdoor $PM_{2.5}$ mass concentrations as well as across levels of total outdoor $PM_{2.5}$. Nitrate in $PM_{2.5}$ was more strongly associated with total $PM_{2.5}$ mass concentrations (r = 0.80) than sulfate (r = 0.74). The correlation between $PM_{2.5}$ and O_x was 0.76.

HRs describing associations between total outdoor $PM_{2.5}$ mass concentrations and mortality with and without adjusting for possible confounding of the version of treatment-outcome relationships by sulfate and nitrate are shown in Tables 2 and 3. For nonaccidental, cardiovascular, and respiratory mortality, HRs for total outdoor $PM_{2.5}$ mass concentrations were confounded by both sulfate and nitrate with inverse associations observed for total outdoor $PM_{2.5}$ when models were adjusted for potential confounding of the version of treatment-outcome relationships. Sulfate and nitrate components were each independently associated with nonaccidental, cardiovascular, and respiratory mortality with stronger associations observed for sulfate. Evidence of confounding was also observed in the relationship between total outdoor $PM_{2.5}$ mass concentrations and lung cancer mortality, but in this case, adjusting for potential confounding of the version of treatment-outcome relationships resulted in elevated risk estimates for total outdoor $PM_{2.5}$ given the inverse associations between both components and lung cancer mortality. HRs for sulfate and nitrate alone (i.e., without total $PM_{2.5}$ in the model) are shown in Supplemental Table 2; http://links.lww.com/EE/A286 and are similar to models including total $PM_{2.5}$ with the exception of lung cancer mortality, where inverse associations were observed for both sulfate and nitrate when $PM_{2.5}$ was included in the model compared with weak/null associations when examined individually.

Table 1.

Descriptive data for total outdoor PM2.5 mass concentrations, nitrate, and sulfate (μg/m³)

Pollutant	Min	1%	5%	10%	25%	Median	75%	90%	95%	99%	Max	Mean	SD
$PM_{2.5}$	2.49	2.89	3.54	4.11	5.58	7.30	9.12	10.33	10.78	11.98	14.47	7.30	2.28
Nitrate	0.00	0.03	0.10	0.18	0.54	0.90	1.22	1.45	1.62	1.93	3.45	0.88	0.47
Sulfate	0.19	0.50	0.62	0.72	0.99	1.56	2.12	2.64	2.96	3.61	5.32	1.62	0.75

Pollutant values reflect 10-year moving averages for each person year.

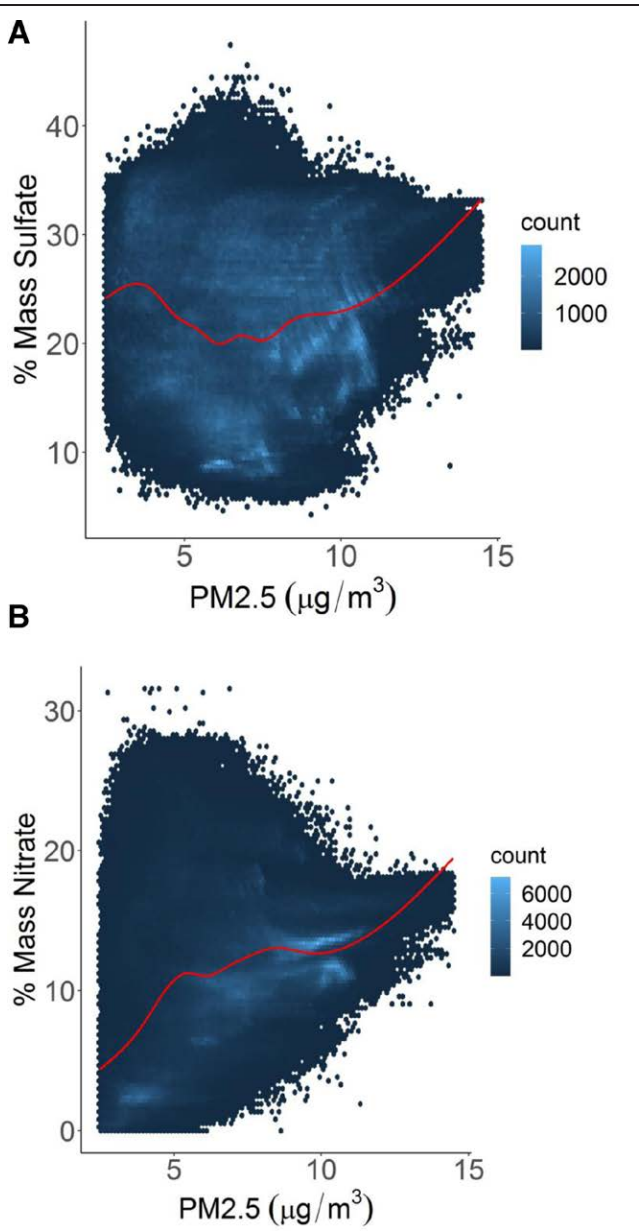

Figure 1. Mass proportions (%) of sulfate (A) and nitrate (B) in outdoor $PM_{2.5}$ across the range of total outdoor $PM_{2.5}$ concentrations. The lines are smoothers demonstrating trends in component mass proportions across the range of total outdoor $PM_{2.5}$ concentrations.

Models including total outdoor $PM_{2.5}$, sulfate, nitrate, and O_x are shown in Supplemental Table 4; http://links.lww.com/EE/A286. For nonaccidental mortality, additionally adjusting for O_x attenuated associations for nitrate, whereas HRs for sulfate remained elevated and the HR for total $PM_{2.5}$ remained largely unchanged. A similar pattern was observed for respiratory and cardiovascular mortality, and none of the lung cancer mortality results were meaningfully changed after adjusting for O_x. Finally, models including outdoor $PM_{2.5}$ as the residuals of linear models regressing outdoor $PM_{2.5}$ concentrations on sulfate, nitrate, or both are shown in Supplemental Table 4; http://links.lww.com/EE/A286. The results of these analyses are consistent (and similar in magnitude) with those above, with positive associations observed between sulfate and nitrate and nonaccidental, cardiovascular, and respiratory mortality and no association (or weak inverse associations) observed for variations in outdoor $PM_{2.5}$ mass concentrations not explained by these two components. Likewise, results for lung cancer mortality were similar to

Table 2.

Hazard ratios (95% confidence intervals) describing the association between total outdoor $PM_{2.5}$ mass concentrations and nonaccidental and cardiovascular mortality with and without adjustment for sulfate and nitrate

Outdoor $PM_{2.5}$ component	Hazard ratio (95% CI)
Nonaccidental mortality	
$PM_{2.5}$ only	
Total $PM_{2.5}$	1.063 (1.053, 1.072)
$PM_{2.5}$ + sulfate	
Total $PM_{2.5}$	0.998 (0.986, 1.010)
Sulfate	1.109 (1.095, 1.122)
$PM_{2.5}$ + nitrate	
Total $PM_{2.5}$	0.996 (0.981, 1.010)
Nitrate	1.072 (1.059, 1.086)
$PM_{2.5}$ + both components	
Total $PM_{2.5}$	0.981 (0.966, 0.995)
Sulfate	1.096 (1.081, 1.111)
Nitrate	1.027 (1.012, 1.041)
Cardiovascular mortality	
$PM_{2.5}$ only	
Total $PM_{2.5}$	1.090 (1.072, 1.110)
$PM_{2.5}$ + sulfate	
Total $PM_{2.5}$	0.983 (0.961, 1.005)
Sulfate	1.183 (1.156, 1.211)
$PM_{2.5}$ + nitrate	
Total $PM_{2.5}$	0.961 (0.935, 0.989)
Nitrate	1.144 (1.117, 1.171)
$PM_{2.5}$ + both components	
Total $PM_{2.5}$	0.937 (0.911, 0.964)
Sulfate	1.146 (1.117, 1.176)
Nitrate	1.074 (1.046, 1.102)

All hazard ratios are scaled by the interquartile range of each pollutant: $PM_{2.5}$ (3.54 μg/m^3), sulfate (1.13 μg/m^3), nitrate (0.675 μg/m^3).

All models were stratified by age (5-year age groups), sex, immigrant status, and census year. Models were adjusted for individual-level covariates derived from census respondents (income quintile, visible minority status, Indigenous identity, educational attainment, labor force status, marital status, and occupation) and ecological covariates assigned using census geographies or postal codes (community size, airshed, urban form, material deprivation, residential instability, dependency, and ethnic concentration).

Table 3.

Hazard ratios (95% confidence intervals) describing the association between total outdoor $PM_{2.5}$ mass concentrations and respiratory and lung cancer mortality with and without adjustment for sulfate and nitrate

Outdoor $PM_{2.5}$ component	Hazard ratio (95% CI)
Respiratory mortality	
$PM_{2.5}$ only	
Total $PM_{2.5}$	1.073 (1.041, 1.106)
$PM_{2.5}$ + sulfate	
Total $PM_{2.5}$	0.981 (0.943, 1.019)
Sulfate	1.162 (1.116, 1.210)
$PM_{2.5}$ + nitrate	
Total $PM_{2.5}$	0.951 (0.906, 0.998)
Nitrate	1.139 (1.093, 1.186)
$PM_{2.5}$ + both components	
Total $PM_{2.5}$	0.934 (0.890, 0.981)
Sulfate	1.122 (1.072, 1.175)
Nitrate	1.077 (1.029, 1.128)
Lung cancer mortality	
$PM_{2.5}$ only	
Total $PM_{2.5}$	1.074 (1.044, 1.104)
$PM_{2.5}$ + sulfate	
Total $PM_{2.5}$	1.110 (1.071, 1.151)
Sulfate	0.947 (0.912, 0.983)
$PM_{2.5}$ + nitrate	
Total $PM_{2.5}$	1.172 (1.120, 1.227)
Nitrate	0.909 (0.981, 1.010)
$PM_{2.5}$ + both components	
Total $PM_{2.5}$	1.175 (1.122, 1.230)
Sulfate	0.985 (0.944, 1.028)
Nitrate	0.916 (0.876, 0.957)

All hazard ratios are scaled by the interquartile range of each pollutant: $PM_{2.5}$ (3.54 μg/m^3), sulfate (1.13 μg/m^3), nitrate (0.675 μg/m^3).

All models were stratified by age (5-year age groups), sex, immigrant status, and census year. Models were adjusted for individual-level covariates derived from census respondents (income quintile, visible minority status, Indigenous identity, educational attainment, labor force status, marital status, and occupation) and ecological covariates assigned using census geographies or postal codes (community size, airshed, urban form, material deprivation, residential instability, dependency, and ethnic concentration).

those above, with inverse associations observed for sulfate and nitrate and positive associations observed for $PM_{2.5}$ mass concentrations not explained by these two components. For lung cancer mortality, a slightly weaker association was observed for residual $PM_{2.5}$ when both sulfate and nitrate were included in the model (HR = 1.118, 95% confidence interval [CI] = 1.080, 1.157) compared with the model including total outdoor PM2.5 mass concentrations (HR = 1.175, 95% CI = 1.122, 1.230).

Concentration-response curves for nonaccidental mortality and sulfate, nitrate, and total outdoor $PM_{2.5}$ mass concentrations adjusted for sulfate or nitrate are shown in Figure 2. Clear positive associations were apparent for both sulfate and nitrate with nonaccidental mortality whereas weaker associations were observed for total outdoor $PM_{2.5}$ mass concentrations with 95% CI generally including the null (or always including the null in the case of models adjusted for sulfate). Concentration-response curves for sulfate and nitrate with cardiovascular and respiratory mortality are shown in Supplemental Figures 2 and 3; http://links.lww.com/EE/A286 and also display clear positive associations.

Discussion

Overall, our findings suggest that studies interested in estimating the health impacts of total outdoor $PM_{2.5}$ mass concentrations (i.e., how much mass is present in the air) need to consider $PM_{2.5}$ composition (i.e., the kind of mass that is present in the air) to avoid bias caused by confounding of version of treatment-outcome relationships. In this study, total outdoor $PM_{2.5}$ mass concentrations were not positively associated with nonaccidental, cardiovascular, or respiratory mortality after adjusting for sulfate or nitrate (potential confounders of version of treatment-outcome relationships), which were each positively associated with all three outcomes (i.e., suggesting that interventions targeting these specific kinds of $PM_{2.5}$ may be more efficient in terms of health benefits than actions targeting total $PM_{2.5}$). Conversely, the strength of the association between total outdoor $PM_{2.5}$ mass concentration and lung cancer mortality increased when sulfate/nitrate was included in the model because of inverse associations between these components and lung cancer mortality. Our analyses including the residuals of linear models regressing total outdoor $PM_{2.5}$ mass concentrations on sulfate and nitrate concentrations support these results, and other studies[5] have also reported decreases in HRs (and in some cases inverse associations similar to those we observed) between total outdoor $PM_{2.5}$ mass concentrations and all-cause, respiratory, and cardiovascular mortality when variability from sulfate and nitrate were removed from total outdoor $PM_{2.5}$. However, Kazemiparkouhi et al[5] reported a decrease in HRs between outdoor $PM_{2.5}$ and lung cancer when variations from sulfate and nitrate were removed, and this contrasts with our findings. Reasons for this discrepancy are not clear, but both studies suggest that the kind of $PM_{2.5}$ mass matters and thus needs to be considered to obtain unbiased estimates of the health risks of total undifferentiated outdoor $PM_{2.5}$ mass concentrations (i.e., the "how much" effect), even if the health impacts of specific $PM_{2.5}$ components are not of primary interest.

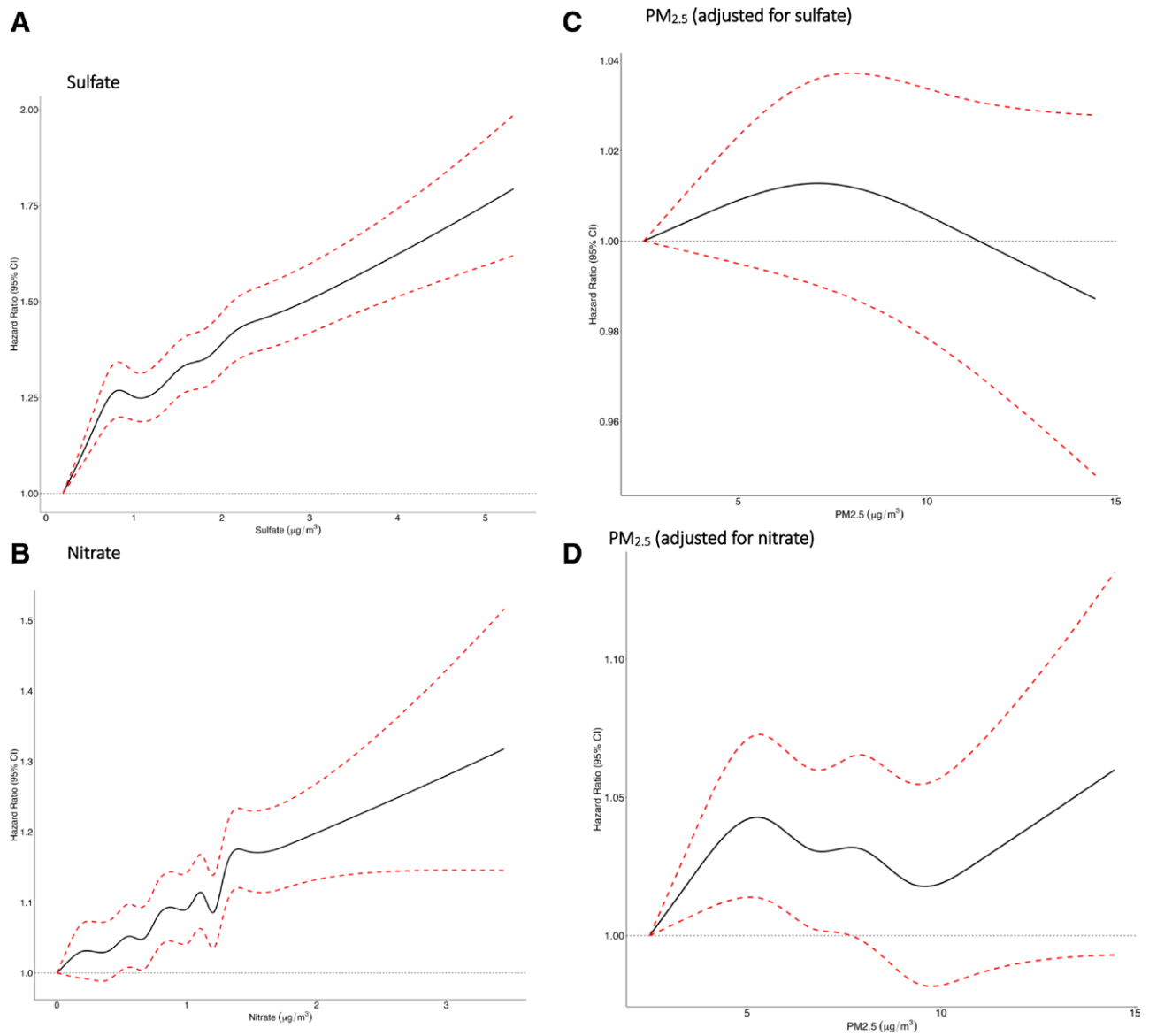

Figure 2. Concentration-response relationships between nonaccidental mortality and sulfate (A), nitrate (B), and total outdoor $PM_{2.5}$ mass concentrations adjusted for sulfate (C) or nitrate (D). Dashed lines indicate 95% confidence intervals. A, Sulfate. B, Nitrate. C, $PM_{2.5}$ (adjusted for sulfate). D, $PM_{2.5}$ (adjusted for nitrate).

Our results raise several interesting questions with respect to possible interventions that may reduce the population health impacts of total outdoor $PM_{2.5}$ mass concentrations. First, because there are always multiple versions of treatment for total outdoor $PM_{2.5}$ mass concentrations, the interpretation of health risk estimates across a gradient of exposure (e.g., between locations with long-term average outdoor concentrations of 5 μg/m^3 vs. 15 μg/m^3) is more subtle than generally appreciated. For example, in the best-case scenario of no confounding, contrasts across levels of total outdoor $PM_{2.5}$ mass concentrations are interpreted as reflecting randomly assigned versions of treatment from the distributions of versions of treatment available in the study population (in our example, this would mean versions of treatment available at 5 μg/m^{3}and 15 μg/m^3). Therefore, in most cases, we are not simply comparing health outcomes across levels of how much mass is present, but we are also comparing different kinds of $PM_{2.5}$ because the composition of total outdoor $PM_{2.5}$ can vary across the range of outdoor $PM_{2.5}$ mass concentrations (as shown for sulfate and nitrate in Figure 2). This makes the planning and implementation of effective and efficient interventions more challenging because the underlying exposures being contrasted are opaque and it is not clear how much of a given health impact for total outdoor $PM_{2.5}$ is attributable to the "how much mass effect" and how much is attributable to the "what kind of mass effect." Moreover, if the underlying components most relevant to the adverse health effects of total outdoor $PM_{2.5}$ mass concentrations are not evenly distributed across space/time, we should not expect homogeneous benefits to health with mandated reductions targeting total outdoor $PM_{2.5}$ mass concentrations.

If we recognize that existing concentration-response curves for outdoor $PM_{2.5}$ typically reflect some combination of the health impacts of how much $PM_{2.5}$ is present in the air and the kind of $PM_{2.5}$ is present, this also raises another interesting thought. Specifically, perhaps there is a continuum across the global $PM_{2.5}$ distribution where the "how much mass effect" is dominant at higher $PM_{2.5}$ concentrations and the "what kind of mass effect" is dominant at lower $PM_{2.5}$ concentrations. In turn, as these effects may have different slopes, the interplay between these two components of the total $PM_{2.5}$ health impact

could contribute to observed heterogeneity in the shapes of concentration-response curves[26] and $PM_{2.5}$-related mortality rates in places with similar total outdoor $PM_{2.5}$ mass concentrations.[27] Further studies are needed to explore this question in depth. In particular, studies of spatial/temporal contrasts in outdoor $PM_{2.5}$ mass concentrations of similar composition in low pollution areas would be helpful in disentangling the "how much mass effect" from the "what kind of mass effect."

One question that may arise from our analysis is why sulfate and nitrate may be independently associated with mortality. For sulfate, one possibility is that increasing sulfate content increases particle acidity, which makes co-occurring transition metals more biologically available.[28] Indeed, previous studies have observed stronger associations between outdoor $PM_{2.5}$ and acute cardiovascular events when both sulfate and metal content are elevated.[29] Moreover, previous studies have observed positive associations between nitrate and biomarkers of inflammation in adults[30] and both components have been independently associated with mortality in the United States.[13] With respect to lung cancer, sulfate, and nitrate might be inversely correlated with the kinds of $PM_{2.5}$ that cause lung cancer, thus explaining the inverse associations observed between lung cancer mortality and these components. Indeed, sulfate and nitrate might just be reasonable surrogates for sources that emit versions of $PM_{2.5}$ that are relevant to health owing to the nature of their specific chemical mixtures. Moreover, it is important to note that these components were selected as examples of potential confounders of the version of treatment-outcome relationships, and there are likely other examples as well. For example, the polycyclic aromatic hydrocarbon (PAH) content of outdoor $PM_{2.5}$ could be a relevant confounder of the version of the treatment-outcome relationship in studies of total outdoor $PM_{2.5}$ and lung cancer because PAH content is related to the version of treatment and PAHs are known human carcinogens.[31] If this is true, the hazard ratio presented for total outdoor $PM_{2.5}$ and lung cancer could be confounded by this source of bias and reflects some unknown combination of the effect of total outdoor $PM_{2.5}$ mass concentration (i.e., how much mass is present) and the effect of the version of treatment (i.e., the kind of mass that is present) on lung cancer mortality.

One point of confusion in our analysis may be related to the framing of the problem in terms of confounding rather than effect modification. We argue that the confounding approach and the multiple versions of the treatment framework are most appropriate for several reasons. First, we reiterate that total outdoor $PM_{2.5}$ mass concentration is the exposure of interest and that in the best-case scenario of no confounding, we are contrasting mortality experiences across randomly assigned versions of treatment from the distributions of versions of treatment available in the study population. Effect modification typically refers to the situation where some third variable modifies the strength of association between the main exposure of interest and the outcome. In the case of effect modification by $PM_{2.5}$ composition, the "third variable" (i.e., variations in $PM_{2.5}$ composition) actually creates systematically different versions of the treatment. In that context, examining the health effects of total $PM_{2.5}$ mass concentrations across strata of varying $PM_{2.5}$ composition is really a comparison of completely different exposures rather than a single version (or randomly assigned version) of the same exposure across levels of an effect modifier.

Particle composition varies across space and time, and if we are interested in estimating the health impacts of total undifferentiated outdoor $PM_{2.5}$ mass concentrations (i.e., the "how much mass effect") and regulating $PM_{2.5}$ in this form, we need to do so in a manner that minimizes potential sources of bias. The fact that $PM_{2.5}$ is a mixture with multiple versions of treatment greatly complicates how we estimate (and interpret) health risks for total outdoor $PM_{2.5}$ mass concentrations. In light of this, we have several options. The first (and likely the easiest) option is to continue with the status quo and ignore the version of treatment/component information recognizing that health risk estimates obtained for total outdoor $PM_{2.5}$ mass concentrations will reflect some unknown combination of the health impacts of how much $PM_{2.5}$ mass is present in the air a given time/location and the kind of $PM_{2.5}$ mass that is present. If we do this, we also need to acknowledge that these health risk estimates do not perfectly align with the technical definition of the current regulatory metric which is focused only on total outdoor $PM_{2.5}$ mass (i.e., how much mass) and not the kind of $PM_{2.5}$ mass. In addition, we should not expect homogeneous health benefits from mandated reductions in total outdoor $PM_{2.5}$ mass concentrations, as the underlying components of health relevance are likely not evenly distributed across space, time, or the distribution of total outdoor $PM_{2.5}$ mass concentrations. The second option is to measure $PM_{2.5}$ components that are likely to act as confounders of the version of treatment-outcome relationships and include these in analyses aimed at quantifying the health impacts of total outdoor $PM_{2.5}$ (i.e., the "how much mass effect"). While we recognize that we do not currently understand all of the components relevant to the health impacts of total outdoor $PM_{2.5}$, existing evidence suggests that some components (such as sulfate and nitrate)[13] are more important than others and thus including these should increase the validity of health risk estimates for total $PM_{2.5}$ mass concentrations.

While this study had many strengths, including a large population-based cohort, national-level exposure information, and updating of exposures for residential mobility it is important to note several limitations. First, as noted above, sulfate and nitrate were selected as two examples of possible confounders of the version of treatment-outcome relationships, and we cannot rule out other components that may contribute to bias in HRs for total $PM_{2.5}$ in a similar manner. In addition, the air pollution exposures used in this study were measured with an error that likely contained components of both classical (e.g., predicted pollutant concentrations at 1-km resolution likely varied around true long-term outdoor values on the same scale) and Berkson-type error (e.g., within each 1-km grid, true outdoor pollutant concentrations for each cohort member are likely distributed around the predicted value for the group). These sources of error likely resulted in a loss of precision, potential bias in health risk estimates for total $PM_{2.5}$ mass concentrations and components (the expectation being toward the null as the classical error component was likely non-differential), and possible residual confounding in health risk estimates for total $PM_{2.5}$ mass concentrations because of measurement error in component estimates used for covariate adjustment. Finally, not everyone who completed the 2006 long-form census was retained in our study population, and people who were living in the territories of Nunavut, Northwest Territories, and Yukon or reporting Inuit or Indigenous identity were less likely to consent to data linkage. This likely impacts the generalizability of our results but is not a serious threat to internal validity (i.e., selection bias), as selection into the study was not systematically related to both exposure and the outcome (or causes of the outcomes given than the outcomes had not yet occurred at the start of follow-up).

In summary, the heterogeneous nature of $PM_{2.5}$ raises several subtle complications related to the estimation and interpretation of the health risks associated with total outdoor $PM_{2.5}$ mass concentrations (i.e., the health effects of how much $PM_{2.5}$ mass is present in the air). In particular, as total outdoor $PM_{2.5}$ varies in composition and toxicity, we need to account for potential confounders of the version of treatment-outcome relationships to avoid bias in health risk estimates for total outdoor $PM_{2.5}$. This can be achieved by considering $PM_{2.5}$ components in our analyses that may be confounders of the version of treatment-outcome relationships or by conducting studies of total outdoor $PM_{2.5}$ across

regions with similar composition. Otherwise, health risk estimates for total outdoor $PM_{2.5}$ will reflect some unknown combination of how much $PM_{2.5}$ mass is present and the kind of $PM_{2.5}$ mass that is present, which is somewhat misaligned with the current regulatory metric, which is focused only on how much $PM_{2.5}$ mass is present in outdoor air.

References

1. Burnett R, Chen H, Szyszkowicz M, et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc Natl Acad Sci USA*. 2018;115:9592–9597.
2. Weichenthal S, Pinault L, Christidis T, et al. How low can you go? air pollution affects mortality at very low levels. *Sci Adv*. 2022;8:eabo3381.
3. Cole SR, Frangakis CE. The consistency statement in causal inference. a definition or an assumption? *Epidemiology*. 2009;20:3–5.
4. Henneman L, Choirat C, Dedoussi I, Dominici F, Roberts J, Zigler C. Mortality risk from United States coal electricity generation. *Science*. 2023;382:941–946.
5. Kazemiparkouhi F, Honda T, Eum KD, Wang B, Manjourides J, Suh HH. The impact of Long-Term PM2.5 constituents and their sources on specific causes of death in a US Medicare cohort. *Environ Int*. 2022;159:106988.
6. Ozkaynak H, Thurston GD. Associations between 1980 U.S. mortality rates and alternative measures of airborne particle concentration. *Risk Anal*. 1987;7:449–461.
7. Rubin DB. Formal modes of statistical inference for causal effects. *J Stat Plan Inference*. 1990;25:279–292.
8. Weichenthal S, Ripley S, Korsiak J. Fine particulate air pollution and the "no-multiple-versions-of-treatment" assumption: does particle composition matter for causal inference? *Am J Epidemiol*. 2023;192:147–153.
9. Chen H, Kaufman JS, Olaniyan T, et al. Changes in exposure to ambient fine particulate matter after relocating and long term survival in Canada: quasi-experimental study. *BMJ*. 2021;375:n2368.
10. Dominici F, Zanobetti A, Schwartz J, Braun D, Sabath B, Wu X. Assessing adverse health effects of long-term exposure to low levels of ambient air pollution: implementation of causal inference methods. *Res Rep Health Eff Inst*. 2022;211:1–56.
11. Wei Y, Yazdi MD, Di Q, et al. Emulating causal dose-response relations between air pollutants and mortality in the medicare population. *Environ Health*. 2021;20:53.
12. van Donkelaar A, Martin RV, Li C, Burnett RT. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with Information from satellites, models, and monitors. *Environ Sci Technol*. 2019;53:2595–2611.
13. Hao H, Wang Y, Zhu Q, et al. National cohort study of long-term exposure to PM2.5 components and mortality in medicare American older adults. *Environ Sci Technol*. 2023;57:6835–6843.
14. Tjepkema M, Christidis T, Bushnik T, Pinault L. Cohort profile: the Canadian Census Health and Environment Cohorts (CanCHECs). *Health Rep*. 2020;30:18–26.
15. Statistics Canada. *Social Data Linkage Environment (SDLE)*. 2017. Available at https://www.statcan.gc.ca/eng/sdle/index. Accessed 28 Jan 2019.
16. Pappin AJ, Christidis T, Pinault L, et al. Examining the shape of the association between low levels of fine particulate matter and mortality across three cycles of the Canadian Census Health and Environment Cohort. *Environ Health Perspect*. 2019;127:107008.
17. Brauer M, Brook JR, Christidis T, et al. Mortality–Air Pollution Associations in Low Exposure Environments (MAPLE): Phase 2. *Res Rep Health Eff Inst*. 2022;2022:1–91.
18. Finès P, Pinault L, Tjepkema M. *Analytical studies: methods and references: imputing postal codes to analyze ecological variables in longitudinal cohorts: exposure to particulate matter in the Canadian Census Health and Environment Cohort database*. Ottawa: Statistics Canada. 2017.
19. Statistics Canada. *Postal Code Conversion File Plus (PCCF+) Version 6D, reference guide: August 2015 postal codes*. Ottawa: Statistics Canada. 2017.
20. Crouse DL, Erickson A, Christidis T, et al. Evaluating the sensitivity of PM2.5-mortality associations to the spatial and temporal scale of exposure assessment at low particle mass concentrations. *Epidemiology*. 2020;31:168–176.
21. Cox, David R. Regression models and life-tables. *J Royal Stat Soc B*. 1972;34:187–220.
22. Crouse DL, Philip S, van Donkelaar A, et al. A new method to jointly estimate the mortality risk of long-term exposure to fine particulate matter and its components. *Sci Rep*. 2016;6:18916.
23. Gordon DL, Janzen M. Suburban nation? Estimating the size of Canada's suburban population. *J Archit Plan Res*. 2013;30:197–220.
24. Matheson FI, Dunn JR, Smith KLW, Moineddin R, Glazier RH. Development of the Canadian Marginalization Index: a new tool for the study of inequality. *Can J Public Health*. 2012;103:S12–S16.
25. VanderWeele TJ, Hernan MA. Causal inference under multiple versions of treatment. *J Causal Inference*. 2013;1:1–20.
26. Chen J, Braun D, Christidis T, et al. Long-term exposure to low-level PM2.5 and mortality: investigation of heterogeneity by harmonizing analyses in large cohort studies in Canada, United States, and Europe. *Environ Health Perspect*. 2023;131:127003.
27. Li X, Jin L, Kan H. Air pollution: a global problem needs local fixes. *Nature*. 2019;570:437–439.
28. Fang T, Guo H, Zeng L, Verma V, Nenes A, Weber RJ. Highly acidic ambient particles, soluble metals, and oxidative potential: a link between sulfate and aerosol toxicity. *Environ Sci Technol*. 2017;51:2611–2620.
29. Weichenthal S, Lavigne E, Traub A, et al. Association of sulfur, transition metals, and the oxidative potential of outdoor PM2.5 with acute cardiovascular events: a case-crossover study of Canadian adults. *Environ Health Perspect*. 2021;129:107005.
30. Wu S, Deng F, Wei H, et al. Chemical constituents of ambient particulate air pollution and biomarkers of inflammation, coagulation and homocysteine in healthy adults: a prospective panel study. *Part Fibre Toxicol*. 2012;9:49.
31. World Health Organization (WHO). *Human health effects of polycyclic aromatic hydrocarbons as ambient air pollutants: report of the Working Group on Polycyclic Aromatic Hydrocarbons of the Joint Task Force on the Health Aspects of Air Pollution*. Copenhagen: WHO Regional Office for Europe; 2021. License: CC BY-NC-SA 3.0 IGO.