

# SCIENTIFIC REPORTS



OPEN

## A New Method to Jointly Estimate the Mortality Risk of Long-Term Exposure to Fine Particulate Matter and its Components

Received: 04 September 2015

Accepted: 30 November 2015

Published: 06 January 2016

Dan L. Crouse<sup>1,†</sup>, Sajeev Philip<sup>2</sup>, Aaron van Donkelaar<sup>2</sup>, Randall V. Martin<sup>2,3</sup>, Barry Jessiman<sup>4</sup>, Paul A. Peters<sup>5</sup>, Scott Weichenthal<sup>6</sup>, Jeffrey R. Brook<sup>7,8</sup>, Bryan Hubbell<sup>9</sup> & Richard T. Burnett<sup>1</sup>

Most studies on the association between exposure to fine particulate matter (PM<sub>2.5</sub>) and mortality have considered only total concentration of PM<sub>2.5</sub> or individual components of PM<sub>2.5</sub>, and not the combined effects of concentration and particulate composition. We sought to develop a method to estimate the risk of death from long-term exposure to PM<sub>2.5</sub> and the distribution of its components, namely: sulphate, nitrate, ammonium, organic mass, black carbon, and mineral dust. We decomposed PM<sub>2.5</sub> exposure into the sum of total concentration and the proportion of each component. We estimated the risk of death due to exposure using a cohort of ~2.4 million Canadians who were followed for vital status over 16 years. Modelling the concentration of PM<sub>2.5</sub> with the distribution of the proportions of components together was a superior predictor for mortality than either total PM<sub>2.5</sub> concentration alone, or all component concentrations modelled together. Our new approach has the advantage of characterizing the toxicity of the atmosphere in its entirety. This is required to fully understand the health benefits associated with strategies to improve air quality that may result in complex changes not only in PM<sub>2.5</sub> concentration, but also in the distribution of particle components.

Ambient fine particulate matter (PM<sub>2.5</sub>) is composed of a complex mixture of solids and liquids (smaller than 2.5 microns in aerodynamic diameter) derived from diverse sources (e.g., human activities, including fossil fuel combustion and industrial activities; and natural sources, including volcanic ash and pollens) that varies in space and time due to atmospheric chemistry, weather, and interactions between it and other pollutants in the atmosphere<sup>1</sup>. The composition of particulate matter, therefore, varies between and within regions of the world, countries, and urban and rural areas, and is influenced by such factors as climate, proximity to an ocean, agricultural activities, transportation activities, and kinds and quantities of point source emitters<sup>2</sup>. The proximity to sources also affects the nature of the mixture of constituents composing PM<sub>2.5</sub> (mixing state) with a greater propensity for external mixtures of particle types closer to sources, and more homogeneity among particle types as an air mass ages and the particles undergo a greater degree of atmospheric processing. The major components of PM<sub>2.5</sub> typically consist of: sulphate; nitrate; ammonium; chloride/sea salt; carbon – described variously as elemental carbon, organic carbon, and black carbon; crustal material, including dust and minerals; and, biological materials and organic mass<sup>2</sup>. The level of toxicity associated with PM<sub>2.5</sub> is strongly affected by its mass and number concentrations, in addition to particle size, shape, chemical composition, and mixing state.

<sup>1</sup>Environmental Health Science and Research Bureau, Health Canada, Ottawa, Ontario, Canada. <sup>2</sup>Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada. <sup>3</sup>Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts. <sup>4</sup>Air Quality Assessment Section, Safe Environments Directorate, Health Canada, Ottawa, Ontario, Canada. <sup>5</sup>Department of Sociology, University of New Brunswick, Fredericton, New Brunswick, Canada. <sup>6</sup>Air Health Science Division, Safe Environments Directorate, Health Canada, Ottawa, Ontario, Canada. <sup>7</sup>Air Quality Research Division, Environment Canada, Downsview, Canada. <sup>8</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. <sup>9</sup>US Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina. <sup>†</sup>Present address: Department of Sociology, University of New Brunswick, Fredericton, New Brunswick, Canada. Correspondence and requests for materials should be addressed to D.L.C. (email: dan.crouse@unb.ca)

The vast majority of epidemiological studies on the association between long-term exposure to PM<sub>2.5</sub> and mortality have considered only total concentrations of PM<sub>2.5</sub>, and not proportions or quantities of its individual components and constituents (e.g.,<sup>3–9</sup>). In this context, it is important to note that while previous studies (including those cohort studies noted above) have found consistently positive associations between long-term exposures to PM<sub>2.5</sub> concentrations and all-cause and cause-specific mortality, the size of effect varied somewhat between places and by cause of death. For example, in their English cohort, Carey *et al.*<sup>4</sup> found stronger associations with respiratory mortality than with cardiovascular mortality, which is inverse to findings in most US-based studies. Additionally, Carey *et al.*<sup>4</sup> and Cesaroni *et al.*<sup>5</sup> (in their Italian cohort) both reported notably weaker associations between long-term exposure to PM<sub>2.5</sub> and all-cause mortality than those reported in Canadian<sup>6</sup>- and US-based studies<sup>7,10</sup>. In addition to differences in methodological choices between these studies (e.g., exposure assessment, covariate adjustment, period of follow-up) and population susceptibility to the effects of exposure, the differences in the findings may be due in part to variations in PM composition.

Given that most previous studies on the long-term effect of exposure have focused primarily on associations between mortality and overall PM<sub>2.5</sub> concentrations, the kinds and sources of particles most responsible for long-term effects on health are not well known. Moreover, as outlined by Kelly and Fussell<sup>1</sup>: “having now firmly established associations between ambient PM and adverse health effects, the biggest gap in our knowledge of PM toxicity relates to which component(s) of ambient PM, and/or which of their physical and chemical characteristic(s), are responsible (p. 505).” There has, however, been recent interest in examining how the distribution of components modifies the effects of total PM<sub>2.5</sub> concentration on mortality risk<sup>11</sup>.

The objective of this study is to develop a mathematical characterization of the combined association between fine particulate matter concentration and the distribution of its components that is suitable for use in survival analysis, which is the method used to estimate risk associated with exposure to long-term concentrations of ambient air pollution and mortality. We illustrate our modelling approach using recently developed estimates of PM<sub>2.5</sub> concentration and the distribution of sulphate, nitrate, ammonium, organic mass, black carbon, and mineral dust for the world<sup>12</sup>. We assigned these estimates to a nationally-representative cohort of ~2.4 million Canadians who completed the 1991 census long form and were linked to mortality records to 2006<sup>6,13,14</sup>.

## Results

Our study included approximately 2.4 million subjects at baseline, nearly 300,000 of whom died of non-accidental causes during the 16 years of follow-up (Table 1). Nearly 60% of subjects lived in the East Central airshed (which encompasses all of southern Ontario and Quebec, including several of Canada’s largest cities). Only 4.9% of subjects self-identified as visible minorities, and 17.1% were immigrants.

At baseline, we found mean exposures to PM<sub>2.5</sub> of 8.3 µg/m<sup>3</sup>. As depicted in Fig. 1, the highest concentrations were located in south-western Ontario, especially in the Windsor-to-Quebec City corridor, and in the western prairies. We present in Supplemental Material, Fig. 1 a map of the individual component concentrations across Canada. Among the six components, organic mass and sulphate contributed the largest proportions of PM<sub>2.5</sub> (e.g., mean proportions at baseline of 0.21 and 0.23, respectively), while mineral dust and black carbon represented the smallest contributions (e.g., mean proportions at baseline of 0.05 each) (see Table 2 for complete distributions).

The Western airshed was characterised by a relatively high proportion of mineral dust (15%) and a low proportion of nitrate (5%). The Prairie airshed was characterised by the highest proportion of nitrate (16%), balanced by the lowest proportions of both black carbon (19%) and sulphate (19%). The West Central airshed had the highest proportion of organic mass (31%), whereas the Southern Atlantic airshed had the lowest proportion of organic mass (20%) and the highest proportion of black carbon (34%). The Northern airshed was composed of only 1% nitrate. The component concentrations are all highly correlated with each other and with total PM<sub>2.5</sub> concentration ( $r > 0.65$ ) except for dust ( $r < 0.50$ ) (Table 3).

We present results for total PM<sub>2.5</sub> concentration and each component concentration individually for non-accidental and cardio-metabolic deaths in Table 4. For each model, the logarithm of total PM<sub>2.5</sub> concentration and the logarithm of component concentration was examined. In addition, we present the Akaike Information Criterion (AIC) as a measure of model fit adjusted for the number of variables in the model. Here, lower AIC values represent improved prediction of mortality.

First, the associations between mortality and total PM<sub>2.5</sub> concentration, and between mortality and each component concentration individually, were positive and statistically significant based on coefficient to standard error ratios greater than two for non-accidental (left-hand models; top half of Table 4) and cardio-metabolic (left-hand models; bottom half of Table 4) mortality; the lone exception being dust. This result suggests that not all of the total PM<sub>2.5</sub> concentration association with mortality can be explained by any single component. Total PM<sub>2.5</sub> concentration, however, was a better predictor of non-accidental mortality compared to any component concentration individually, as measured by the AIC (Table 4).

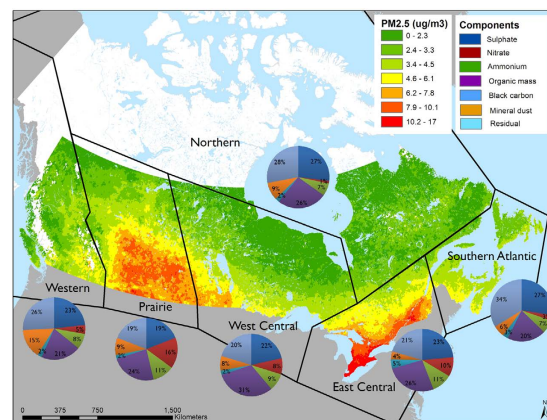
Next, models consisting of all six components examined simultaneously (middle models of Table 4) were better predictors of mortality (both causes of death) than was total PM<sub>2.5</sub> alone. This finding suggests that understanding the complete composition of the atmosphere improves mortality prediction. The coefficients in the multi-component concentration models were highly unstable compared to those in the single component concentration models, due likely to the very high correlations between component concentrations (Table 3). The high correlations are due, in part, to the manner in which these component concentrations were calculated as the product of total concentration times the component proportion. Thus, in areas of high or low overall PM<sub>2.5</sub> concentration, the component concentration has a tendency to also be relatively high or low.

Third, we developed a methodology that somewhat elevates this confounding issue by decomposing total PM<sub>2.5</sub> concentration into its component proportions. Here, the joint total concentration plus the distribution of component proportions model (right-hand models, Table 4) was a better predictor than was PM<sub>2.5</sub> concentration alone, each of the component concentrations separately, and all six together jointly for both non-accidental

Status at baseline	Subjects		Deaths		Hazard Ratio <sup>b</sup>	95% CI	
	n <sup>a</sup>	%	n <sup>a</sup>	%			
All	2,451,855	100	299,165	100	1.012	1.010	1.014
Sex							
men	1,221,215	49.8	182,690	61.1	-	-	-
women	1,230,640	50.2	116,475	38.9	-	-	-
Age							
25–34	679,580	27.7	6,150	2.1	-	-	-
35–64	1,398,460	57.0	119,125	39.8	-	-	-
65–74	248,645	10.1	112,280	37.5	-	-	-
75–90	125,170	5.1	61,610	20.6	-	-	-
Aboriginal ancestry							
yes	100,230	4.1	7,475	2.5	1.165	1.010	1.014
no	2,351,625	95.9	291,280	97.5	-	-	-
Visible minority status							
yes	120,450	4.9	7,885	2.6	0.815	0.796	0.835
no	2,331,405	95.1	291,280	97.4	-	-	-
Marital status							
divorced/separated/widowed	321,705	13.1	74,400	24.9	1.221	1.209	1.233
married/common law	1,832,870	74.8	198,920	66.5	-	-	-
single	297,280	12.1	25,845	8.6	1.276	1.259	1.293
Income quintiles							
lowest	382,190	15.6	48,225	16.1	1.299	1.282	1.316
lower middle	466,625	19.0	71,235	23.8	1.192	1.177	1.207
middle	516,300	21.1	74,400	24.9	1.126	1.112	1.140
upper middle	538,430	22.0	56,130	18.8	1.063	1.049	1.076
upper	548,310	22.4	49,175	16.4	-	-	-
Education							
<high school graduation	846,875	34.5	170,675	57.1	1.417	1.393	1.442
high school graduation with or without trade certificate	900,080	36.7	84,575	28.3	1.277	1.255	1.299
some postsecondary, or college diploma	383,075	15.6	24,985	8.4	1.142	1.120	1.165
>=university degree	321,825	13.1	18,930	6.3	-	-	-
Occupational class							
not applicable	587,970	24.0	187,685	62.7	1.291	1.259	1.324
management	204,480	8.3	12,455	4.2	1.038	1.011	1.065
professional	279,465	11.4	11,485	3.8	-	-	-
technical	596,375	24.3	37,745	12.6	1.091	1.066	1.116
semi-skilled	596,590	24.3	34,820	11.6	1.145	1.119	1.172
unskilled	187,075	7.6	14,975	5.0	1.224	1.192	1.257
Labour force status							
employed	1,607,040	65.5	86,790	29.0	-	-	-
unemployed	162,195	6.6	9,545	3.2	1.173	1.148	1.199
not in labour force	682,195	27.8	202,830	67.8	1.280	1.258	1.301
Immigrant status							
Canadian born	2,031,895	82.9	241,190	80.6	1.191	1.179	1.203
immigrant	419,960	17.1	57,975	19.4	-	-	-
Airshed							
Western	328,055	13.4	40,750	13.6	-	-	-
Prairie	330,270	13.5	33,405	11.2	-	-	-
West Central	98,770	4.0	12,735	4.3	-	-	-
East Central	1,448,740	59.1	180,530	60.3	-	-	-
South Atlantic	239,360	9.8	30,880	10.3	-	-	-
Northern	6,660	0.3	865	0.3	-	-	-

**Table 1. Descriptive statistics of cohort subjects at baseline and fully-adjusted hazard ratios for risk factors included in the survival models for all non-accidental causes of death.** <sup>a</sup>subject counts rounded to nearest 5; percentages based on original values. <sup>b</sup>from models adjusted for the following personal covariates: PM<sub>2.5</sub>, aboriginal ancestry, visible minority status, highest level of education, employment status, occupational class,

immigrant status, marital status, and income quintile; and the following contextual covariates: CD and CT-CD % of immigrants, % of adults without high school diploma, % of subjects in lowest income quintile; models stratified by age, sex, and airshed; hazard ratios per  $\mu\text{g}/\text{m}^3$  increment in exposure to  $\text{PM}_{2.5}$ .



**Figure 1.** Median concentrations of  $\text{PM}_{2.5}$  (2001–2010) and relative component proportions by airshed. Map created in ArcGIS Desktop 10.2. ESRI, Redlands, CA.

	$\text{PM}_{2.5}$ ( $\mu\text{g}/\text{m}^3$ )	Components: proportions					
		Sulphate	Nitrate	Ammonium	Particulate organic mass	Black carbon	Mineral dust
Mean	8.26	0.21	0.12	0.10	0.23	0.05	0.05
Max	17.00	0.38	0.38	0.16	0.39	0.07	0.20
95%	13.90	0.29	0.29	0.14	0.29	0.07	0.10
90%	12.60	0.25	0.25	0.13	0.28	0.06	0.08
75%	10.80	0.22	0.22	0.11	0.27	0.06	0.07
Median	8.00	0.20	0.20	0.11	0.23	0.06	0.03
25%	5.60	0.19	0.19	0.10	0.19	0.03	0.03
10%	4.20	0.18	0.18	0.08	0.15	0.02	0.02
5%	3.60	0.18	0.18	0.07	0.13	0.02	0.02
Min	1.20	0.14	0.14	0.02	0.03	0.00	0.02

**Table 2.** Distributions of exposures at baseline.

	$\text{PM}_{2.5}$	Sulphate	Nitrate	Ammonium	Particulate organic mass	Black carbon	Mineral dust
$\text{PM}_{2.5}$	1	0.93	0.89	0.96	0.81	0.91	0.42
Sulphate		1	0.75	0.90	0.72	0.80	0.47
Nitrate			1	0.93	0.66	0.89	0.23
Ammonium				1	0.76	0.90	0.40
Particulate organic mass					1	0.71	0.49
Black carbon						1	0.30
Mineral dust							1

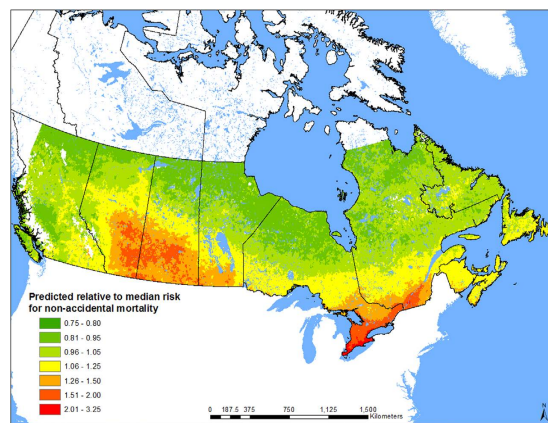
**Table 3.** Pearson correlations between concentrations ( $\mu\text{g}/\text{m}^3$ ) of  $\text{PM}_{2.5}$  and the six components, as assigned to subjects at baseline and stratified by airshed.

(AIC = 5,731,809.8) and cardio-metabolic (AIC = 2,207,515.9) mortality. Additionally, we calculated likelihood ratio tests comparing models with  $\text{PM}_{2.5}$  alone and those with  $\text{PM}_{2.5}$  plus component proportions for models with both causes of death. Here we found for deaths from non-accidental and cardio-metabolic causes chi-square values of 42.6 and 48.8, respectively, with six degrees of freedom that resulted in p-values < 0.0001 in both cases. The evidence from these AIC and likelihood ratio tests support modelling total  $\text{PM}_{2.5}$  concentration and the distribution of its components together.

We present in Fig. 2 the spatial distribution of predicted, relative to median risk estimates, for non-accidental mortality in each  $10 \times 10$  km grid cell across Canada. As would be expected given the results of the models described earlier, these spatial patterns of risk largely mirror the patterns of  $\text{PM}_{2.5}$  concentrations as presented in Fig. 1. One notable difference between the patterns depicted in these two figures is the relatively moderate

Cause of Death	Pollutant	Single-component concentration models			Multi-component concentration models			PM <sub>2.5</sub> + component distribution models		
		AIC	Estimate	Std. Error	AIC	Estimate	Std. Error	AIC	Estimate	Std. Error
Non-accidental		(5,731,800+)			(5,731,800+)			(5,731,800+)		
					25.4			9.8		
	PM <sub>2.5</sub>	40.5	0.084	0.009					0.080	0.012
	Sulphate	43.9	0.092	0.010		0.127	0.022		0.025	0.043
	Nitrate	84.7	0.030	0.004		0.004	0.012		-0.039	0.023
	Ammonium	76.0	0.049	0.006		0.021	0.028		0.033	0.027
	Organic mass	117.5	0.031	0.007		-0.032	0.015		-0.053	0.020
	Black carbon	83.9	0.037	0.005		-0.009	0.014		-0.049	0.025
	Dust	135.7	0.010	0.009		-0.042	0.012		-0.068	0.020
Cardio-metabolic		(2,207,500+)			(2,207,500+)			(2,207,500+)		
					22.6			15.9		
	PM <sub>2.5</sub>	52.7	0.133	0.014					0.143	0.02
	Sulphate	46.5	0.153	0.015		0.235	0.036		0.177	0.068
	Nitrate	87.9	0.051	0.007		0.034	0.019		0.025	0.037
	Ammonium	76.0	0.085	0.010		0.024	0.045		0.054	0.043
	Organic mass	112.8	0.065	0.011		0.014	0.024		0.017	0.032
	Black carbon	100.8	0.055	0.008		-0.091	0.023		-0.086	0.041
	Dust	143.9	0.023	0.014		-0.085	0.019		-0.077	0.032

**Table 4. Results from single and multi-component models for deaths from non-accidental causes (n = 299,165) and cardio-metabolic causes (n = 116,725).**

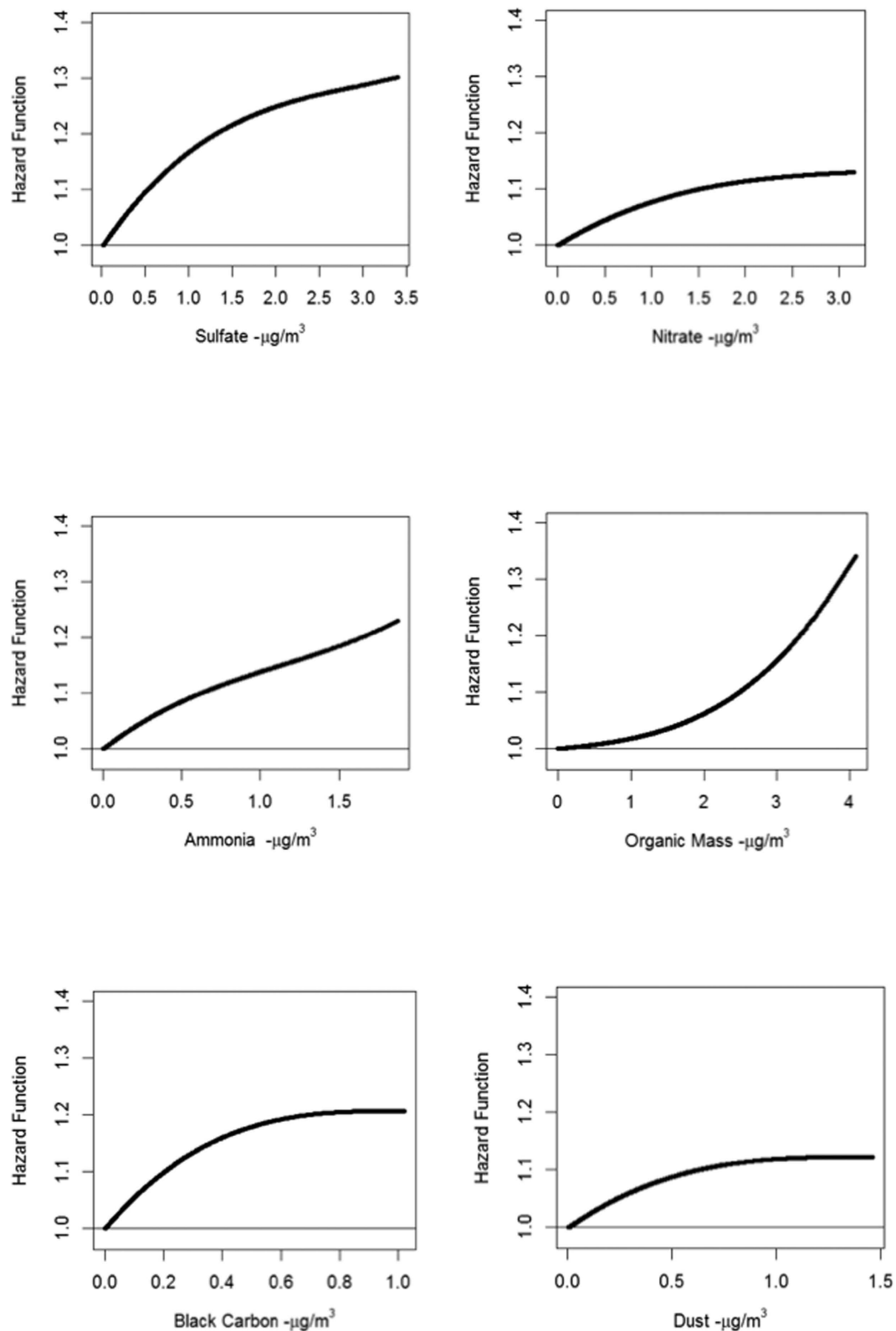


**Figure 2. Predicted relative to median risk for non-accidental mortality.** Map created in ArcGIS Desktop 10.2. ESRI, Redlands, CA.

estimates of risk throughout the maritime provinces of eastern Canada, compared to the relatively low estimates of PM<sub>2.5</sub> in these areas. The somewhat elevated levels of risk in these provinces may be due to the relatively high proportions of sulphate and black carbon of the PM<sub>2.5</sub> in that region (as noted in the pie charts of Fig. 1).

We note that the influence of any single component proportion cannot be evaluated in isolation of the other components. Each component is only assessed appropriately by considering the joint change in concentration and the component distribution together, as can be done using chemical transport models. Here, specific changes in pollutant sources, for example, can be modelled simultaneously.

We present in Fig. 3 the smoothing spline representation of our predicted hazard function for non-accidental mortality vs. each individual component concentration. Here we can see that the risk of mortality is positively associated with each component concentration, possibly due to the role total PM<sub>2.5</sub> concentration plays in both our hazard function and component concentrations. We also observed, however, that the range in predictions of the smooth representation of our hazard function is greatest for sulphate and organic mass, moderate for ammonia and black carbon, and least for nitrate and dust. These observations suggest an ordering of the influence of the individual components on our predicted hazard function.



**Figure 3.** Smoothed plots of predicted risk of non-accidental mortality and component concentrations.

## Discussion

We found that an estimate of total  $\text{PM}_{2.5}$  concentration was a better predictor of mortality than was an estimate of exposure to any individual component of  $\text{PM}_{2.5}$ . Moreover, we found our best predictions of mortality for models with both the total concentration of  $\text{PM}_{2.5}$ , along with the relative proportions of the six components, together.

Knowledge of the relative proportions of the components that make up the total concentrations of particulate matter improved our ability to predict mortality.

A host of studies conducted around the world have now demonstrated that long-term exposures to PM<sub>2.5</sub> are associated with increased mortality<sup>3–7,9,10,15,16</sup>. Moreover, several studies have attempted to identify the constituents or components of PM<sub>2.5</sub> that may be most responsible for PM<sub>2.5</sub>-related health outcomes<sup>17–23</sup>.

In this paper, we took a different approach in that we examined how the association between PM<sub>2.5</sub> and mortality is modified by the distribution of components at a given geographic location. In particular, we developed a new mathematical model to decompose PM<sub>2.5</sub> concentrations into total concentration and the distribution of components, suitable for use in standard survival models.

Initially we examined the association between the concentration of each component and mortality, where component concentration was constructed as the product of total PM<sub>2.5</sub> and component proportions. Because each component concentration is a function of total PM<sub>2.5</sub> concentration, the component concentrations were highly correlated, thus making it difficult to estimate mortality associations for each component simultaneously.

To address this issue, we decomposed PM<sub>2.5</sub> into total concentration plus the distribution of component proportions and jointly estimated mortality associations for these new variables. We also demonstrated that such decomposition improved our ability to predict mortality compared to PM<sub>2.5</sub> concentration alone or by simultaneously modelling all component concentrations together. This is an important advance as the impetus behind the assessment of PM<sub>2.5</sub> components is driven largely by the assumption that knowledge of individual constituents may allow for more efficient regulation if the most important components of PM<sub>2.5</sub> can be targeted for reduction.

A key strength of our study is the large size of our nationally-representative cohort. CanCHEC is one of the largest cohort studies in the world, and includes data on many potential risk factors for mortality. Our exposure models covered all but the largely unpopulated northern regions of the country. As such, we were able to include subjects from every province and from urban and rural areas across Canada.

A noted limitation of our cohort dataset is that personal covariates were available only at baseline. Subjects were followed for vital status, but not for socio-demographic information (e.g., income, employment status). A related limitation of our analysis is that our exposure models described a fixed period of time (i.e., mean concentrations of PM<sub>2.5</sub> over the period 2001–2010), and therefore may not accurately describe longer-term patterns of pollutant concentrations. It is also important to note that other components of PM<sub>2.5</sub>, such as metals, were not examined in this study and may play an important role in overall PM<sub>2.5</sub> toxicity owing to their ability to cause oxidative stress<sup>24,25</sup>. Uncertainty in the PM<sub>2.5</sub> components may also impact the resultant model, particularly where one component is better represented than another. Estimates of sulphate, for example, show particularly good agreement with ground observations ( $r = 0.95$ ) suggesting associations with this component will be better captured than other components. If possible, future studies should expand this kind of analysis to include other potentially important particle constituents as such data become available.

More generally, the method described in this study can be used to estimate the population health burden of disease associated with specific strategies to improve air quality. Such strategies may alter the components in complex ways. We suggest risk models consisting of multiple components estimated simultaneously along with total concentration be used over models with either concentration alone, or component concentrations individually or even simultaneously. This approach provides policy analysts with a powerful tool given that estimates of changes in the atmospheric mix of pollution can be made throughout the world. The associations between mortality and PM<sub>2.5</sub> and its composition, however, may differ in other parts of the world where populations have differences in susceptibility to the effects of exposure, and where overall concentrations of PM<sub>2.5</sub> are much higher (as effects may not be linear). We recommend that similar analyses be conducted with cohort studies of ambient air pollution throughout the world in order to obtain more stable and generalizable estimates of mortality risk associated with exposure to the atmospheric mixtures of particle phase pollution.

## Methods

**The study cohort.** The Canadian Census Health and Environment Cohort (CanCHEC) has been described in detail previously<sup>6,13,14</sup>. CanCHEC is a population-based cohort of subjects who were 25 years of age and older at baseline; a usual resident of Canada on the Census reference day (June 4, 1991); not a resident of an institution such as a prison, hospital, or nursing home; and among the 20% of Canadian households selected for enumeration with the long-form Census questionnaire. Subjects in CanCHEC were linked to the Canadian Mortality Database using deterministic and probabilistic linkage methods from June 4, 1991 through December 31, 2006<sup>13</sup>. The date of death and the underlying cause of death were extracted from death certificates coded to the International Classification of Diseases, 9th Revision (ICD-9)<sup>26</sup> for deaths before 2000, and to ICD-10<sup>27</sup> for those that occurred from 2000 onward. We also linked annual place of residence (six-character postal code) using Historical Tax Summary Files from 1990 through 2006<sup>13</sup>. In urban areas, six-character residential postal codes correspond to one side of a city block or to a single apartment building; in rural areas they may represent a larger area. We excluded recent immigrants (those immigrating in the eight years prior to baseline) given that they tend to have notably different health behaviours and outcomes<sup>28</sup> and may have notably different past exposures, than the native-born Canadian population.

**Assignment of exposure to ambient air pollution.** We assigned estimates of concentrations of PM<sub>2.5</sub> and the relative proportions of several constituent components to the representative point of each subject's residential postal code for every available year between 1990 and 2006 (i.e., beginning one year prior to baseline). A subject's exposure was coded as missing in years for which no residential postal code was available, which could indicate that the subject had left Canada, moved to an institution, or that they had not filed an income tax return that year.

We derived estimates of  $PM_{2.5}$  from observations from three satellite instruments (MISR, MODIS, and SeaWiFS) to represent annual concentrations during the period 2001–2010 on a grid with a spatial resolution of approximately  $10\text{ km} \times 10\text{ km}$  and included coverage up to  $60^\circ\text{N}$  (i.e., excluding the northern regions of Canada)<sup>29</sup>. These  $PM_{2.5}$  estimates were generated by relating satellite retrievals of aerosol optical depth (AOD) to near-surface  $PM_{2.5}$  using the AOD to  $PM_{2.5}$  relationship calculated from the GEOS-Chem chemical transport model. Significant agreement was found with North American ground-based monitors ( $r = 0.76$ ). Inter-annual variation in  $PM_{2.5}$  between 2001 and 2010 was inferred from two radiometrically stable satellite instruments (MISR and SeaWiFS) following the same methodology<sup>30</sup>.

We estimated the proportion of  $PM_{2.5}$  attributed to six components, namely: sulphate, nitrate, ammonium, organic mass, black carbon, and mineral dust. The proportion attributed to each component was calculated with the GEOS-Chem chemical transport model as described in Philip *et al.*<sup>12</sup> by linearly interpolating the native resolution of  $50\text{ km}$  by  $66.7\text{ km}$  grid resolution to the  $10\text{ km}$  by  $10\text{ km}$  of the gridded satellite observations and applying the distribution of components to the estimated  $PM_{2.5}$  concentration. We applied a species-dependent sampling correction and an additional correction/capping to avoid unrealistic retrievals. Evaluation with ground-based observations across the United States and Canada indicated significant agreement for all components. The remaining percentage was assigned to a residual component such that percentages summed to unity.

For each subject and for each year of follow-up we assigned a single-year lagged estimate of exposure to  $PM_{2.5}$  and the proportion of each component. This method of exposure assignment allowed us to incorporate into our models variability in exposure associated with annual residential mobility patterns. The spatial structures of the exposure surfaces were not assessed in a time-dependent manner and thus variation in exposure for each subject was attributed solely to residential mobility.

**Decomposition into distribution of particulate components.** We propose to estimate the joint association between mortality and long-term exposure to concentrations of fine particulate matter, accompanied by the distribution of components. That is, to consider simultaneously the total concentration of  $PM_{2.5}$  and its makeup. In an earlier study with this cohort, we observed the logarithm of  $PM_{2.5}$  to be a stronger predictor of mortality than  $PM_{2.5}$  itself<sup>6</sup>. We therefore develop our approach to modelling  $PM_{2.5}$  and the component proportions based on the logarithm of  $PM_{2.5}$  concentration.

To do this, we set up the following notation. Let  $m$  denote the  $PM_{2.5}$  concentration in  $\mu\text{g}/\text{m}^3$  and let  $p = (p_1, \dots, p_C)'$  be a vector of proportions of total concentration attributable to the  $C$  components of interest with  $p_{C+1}$  denoting the proportion of the residual component. Total  $PM_{2.5}$  concentration can then be decomposed into the concentration times a product of the component proportions normalized to unity by its geometric mean in the following manner:

$$m = m \times \prod_{c=1}^{C+1} \left( p_c / \prod_{c=1}^C p_c^{1/(C+1)} \right) \equiv m \times \prod_{c=1}^{C+1} \tilde{p}_c \quad (1)$$

with  $\prod_{c=1}^{C+1} p_c^{1/(C+1)}$  the geometric mean of the proportions, since  $\prod_{c=1}^{C+1} \tilde{p}_c = 1$  by construction. Here,

$$\tilde{p}_c = \frac{p_c}{\prod_{c=1}^{C+1} p_c^{1/(C+1)}} \quad (2)$$

is the ratio of the  $c^{\text{th}}$  component proportion to its geometric mean. We then have

$$\log(m) = \log(m) + \sum_{c=1}^{C+1} \log(\tilde{p}_c). \quad (3)$$

The corresponding part of the hazard function that involves fine particulate matter and its components in the Cox proportional hazards model is

$$HR(m|p) = \exp \left\{ \theta \times \log(m) + \sum_{c=1}^C \theta_c \times \log(\tilde{p}_c) \right\} = HR(m) \times \prod_{c=1}^C \tilde{p}_c^{\theta_c}, \quad (4)$$

with  $HR(m) = \exp\{\theta \times \log(m)\} = m^\theta$  the hazard ratio model for total concentration in a model with the component proportions. Here  $\theta$  is an unknown risk parameter associated with total concentration and the  $\theta_1, \dots, \theta_C$  are the corresponding unknown parameters for the  $C$  components, which relate the deviation in risk for each component to the risk of total concentration on the logarithmic scale. This approach allows us to modulate the hazard ratio based on spatial variations in component proportions.

Note that the residual component is not included in the survival model due to over specification, since, given  $p_1, \dots, p_C, p_{C+1}$  is known. Estimates of the unknown regression parameters are obtained by fitting a Cox proportional hazards regression model with all the mortality risk factors included in Table 1,  $PM_{2.5}$  concentration ( $m$ ), and the normalized component proportions ( $\tilde{p}_1, \dots, \tilde{p}_C$ ).

The manner in which the components modulate the hazard ratio is complex, with an increase in any single component proportion requiring decreases in the others, since all component proportions must sum to unity. Note that for component proportions greater than the geometric mean of all component proportions  $\tilde{p}_c > 1$ , and if a proportion is less than the geometric mean of all component proportions then  $\tilde{p}_c < 1$ . To interpret our model we need to consider four cases:



1. If  $\theta_c > 0$  and  $\tilde{p}_c > 1$ ,  $HR(m|p_c) > HR(m)$  and will increase as  $\tilde{p}_c$  increases,
2. If  $\theta_c < 0$  and  $\tilde{p}_c > 1$ ,  $HR(m|p_c) < HR(m)$  and will decrease as  $\tilde{p}_c$  increases,
3. If  $\theta_c > 0$  and  $\tilde{p}_c < 1$ ,  $HR(m|p_c) < HR(m)$  and will increase as  $\tilde{p}_c$  increase.,
4. If  $\theta_c < 0$  and  $\tilde{p}_c < 1$ ,  $HR(m|p_c) > HR(m)$  and will decrease as  $\tilde{p}_c$  increases,

where  $HR(m|p_c) = HR(m) \times \tilde{p}_c^{\theta_c}$  is the hazard ratio for concentration and a single marginal component within the joint model. (Note: We include as Supplemental Information the method and formulas for deriving the  $PM_{2.5}$  and distribution of components decomposed for linear hazard models).

**Main statistical analyses.** We used Cox proportional hazards models to estimate the associations between exposure estimates and mortality. We estimated hazard ratios (HRs) stratified by sex, five-year age groups from age 25 to 89, and by regional airshed. We restricted our study to subjects less than 90 years of age due to potential inaccuracies in record-linkages among older subjects (e.g., the address reported on the annual income tax filings of older subjects may reflect those of next of kin, or of institutional facilities). A collaborative approach known as the Air Quality Management System (AQMS) has been developed recently by Canadian federal, provincial, and territorial governments for reducing air pollution in Canada. Among other tasks, the AQMS has delineated six regional airsheds that cut across jurisdictional boundaries in Canada and that have similar air quality characteristics or air movement patterns (Personal communication with Jeff Brook, Environment Canada, December 2014). We stratified by airshed with the intention of explaining broad spatial variation in both mortality and air pollution while still maintaining some variability in exposure within each airshed.

We adjusted our models for aboriginal ancestry, visible minority status, immigrant status, marital status, highest level of education, employment status, occupational classification, and quintiles of household income (see Table 1 for coding). We also calculated time-varying contextual variables adjusted for regional variations across Canada (i.e., census division (CD) means subtracted from census-tract (CT) means) describing the proportion of unemployed adults, the proportion of adults who had not completed high school, and the proportion of individuals in the lowest income quintile. Census tracts correspond roughly to the size of a neighbourhood and census divisions correspond roughly to the size of a city or county. These time-varying contextual covariates describe characteristics of subjects' current residential location during each year of follow-up.

We developed models for two causes of mortality, namely all non-accidental causes (ICD-9 codes  $< 800$ ; ICD-10 codes starting with letters A through R) and cardio-metabolic diseases (i.e., circulatory plus diabetes; (ICD-9 codes: 390–459 and 250; ICD-10 codes: I00–I99 and E10–E14)). Subjects were censored at time of death or if they were lost to follow-up due to end of study period or lack of postal code information. All analyses were conducted in SAS Version 9.3.

We also estimated the association between the logarithm of each component concentration individually (i.e.,  $\log(PM_{2.5} \times \text{component proportion})$ ) and mortality for the purpose of presenting results in a manner used by previous investigators. In addition, we present results based on all six component concentrations together in the same survival model.

We calculated Pearson correlations ( $r$ ) between the component concentrations and untransformed  $PM_{2.5}$  concentration as assigned to subjects at baseline stratified by airshed. Here, we calculated the correlations by airshed separately and then determined a weighted average of the airshed-specific correlations with weights proportional to the number of subjects in each airshed.

Lastly, we calculated the predicted risk estimate in each  $10 \times 10$  km grid cell ( $n = 75,118$  cells) across Canada based on the local concentration of  $PM_{2.5}$  and relative proportions of the components. We then estimated the relationship, using smoothing splines, between the predicted hazard function for non-accidental mortality and each component concentration in order to understand how our complex hazard model related to the more standard representation of component information, namely each component concentration.

**Approval.** The Canadian Census Health and Environment Cohort (CanCHEC) received approval by the Statistics Canada Policy Committee (reference number 012-2001) after consultation with the Statistics Canada Confidentiality and Legislation Committee, Data Access and Control Services Division, and the Federal Privacy Commissioner. This approval is equivalent to that of standard research ethics boards.

## References

1. Kelly, F. J. & Fussell, J. C. Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmos. Environ.* **60**, 504–526 (2012).
2. Harrison, R. M. & Yin, J. Particulate matter in the atmosphere: Which particle properties are important for its effects on health? *Sci. Total Environ.* **249**, 85–101 (2000).
3. Beelen, R. *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* **383**, 785–795 (2014).
4. Carey, I. M. *et al.* Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. *Am. J. Respir. Crit. Care Med.* **187**, 1226–1233 (2013).
5. Cesaroni, G. *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.* **121**, 324–331 (2013).
6. Crouse, D. L. *et al.* Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. *Environ. Health Perspect.* **120**, 708–714 (2012).
7. Laden, F., Schwartz, J., Speizer, F. E. & Dockery, D. W. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities Study. *Am. J. Respir. Crit. Care Med.* **173**, 667–672 (2006).
8. Eftim, S., Samet, J., Janes, H., McDermott, A. & Dominici, F. Fine particulate matter and mortality: A comparison of the six cities and American cancer society cohorts with a Medicare cohort. *Epidemiology* **19**, 209–216 (2008).
9. Krewski, D. *et al.* Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. *Res. Rep. Health. Eff. Inst.* 5–114 (2009).

10. Puett, R. C. *et al.* Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *Environ. Health Perspect.* **117**, 1697–1701 (2009).
11. Chung, Y., Dominici, F., Wang, Y., Coull, B. & Bell, M. Associations between Long-Term Exposure to Chemical Constituents of Fine Particulate Matter (PM<sub>2.5</sub>) and Mortality in Medicare Enrollees in the Eastern United States. *Env. Heal. Perspect.* **123**, 467–474 (2015).
12. Philip, S. *et al.* Global Chemical Composition of Ambient Fine Particulate Matter for Exposure Assessment. *Environ. Sci. Technol.* **48**, 13060–13068 (2014).
13. Peters, P. a. *et al.* Data resource profile: 1991 Canadian census cohort. *Int. J. Epidemiol.* **42**, 1319–1326 (2013).
14. Crouse, D. L. *et al.* Within- and between-city contrasts in nitrogen dioxide and mortality in 10 Canadian cities; a subset of the Canadian Census Health and Environment Cohort (CanCHEC). *J. Expo. Sci. Environ. Epidemiol.* 1–8 (2015). doi: 10.1038/jes.2014.89
15. Jerrett, M. *et al.* Spatial analysis of air pollution and mortality in California. *Am. J. Respir. Crit. Care Med.* **188**, 593–599 (2013).
16. 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.* **120**, 965–970 (2012).
17. Laden, F., Neas, L. M., Dockery, D. W. & Schwartz, J. Association of fine particulate matter from different sources with daily mortality in six U.S. cities. *Environ. Health Perspect.* **108**, 941–947 (2000).
18. Mostofsky, E. *et al.* Modeling the association between particle constituents of air pollution and health outcomes. *Am. J. Epidemiol.* **176**, 317–326 (2012).
19. Ostro, B. *et al.* Long-term exposure to constituents of fine particulate air pollution and mortality: Results from the California Teachers Study. *Environ. Health Perspect.* **118**, 363–369 (2010).
20. Son, J. Y., Lee, J. T., Kim, K. H., Jung, K. & Bell, M. L. Characterization of fine particulate matter and associations between particulate chemical constituents and mortality in Seoul, Korea. *Environ. Health Perspect.* **120**, 872–878 (2012).
21. Beelen, R. *et al.* Review Natural-Cause Mortality and Long-Term Exposure to Particle Components : An Analysis of 19 European Cohorts within the Multi-Center ESCAPE Project. *Env. Heal. Perspect* **123**, 525–533 (2015).
22. Zanobetti, A., Franklin, M., Koutrakis, P. & Schwartz, J. Fine particulate air pollution and its components in association with cause-specific emergency admissions. *Env. Heal.* **8**, (2009).
23. Franklin, M., Koutrakis, P. & Schwartz, P. The role of particle composition on the association between PM<sub>2.5</sub> and mortality. *Epidemiology* **19**, 209–216 (2008).
24. Donaldson, K. *et al.* Combustion-derived nanoparticles: A review of their toxicology following inhalation exposure. *Part. Fibre Toxicol.* **2**, (2005).
25. Künzli, N. *et al.* Comparison of oxidative properties, light absorbance, and total and elemental mass concentration of ambient PM<sub>2.5</sub> collected at 20 European sites. *Environ. Health Perspect.* **114**, 684–690 (2006).
26. WHO. *Manual of the international statistical classification of diseases, injuries and causes of death, ninth revision.* (World Health Organization, 1977).
27. WHO. *International statistical classification of diseases and related health problems. tenth revision.* (World Health Organization, 1992).
28. Ng, E. The healthy immigrant effect and mortality rates. *Health Rep.* **22**, 25–9 (2011).
29. Van Donkelaar, A., Martin, R. V., Brauer, M. & Boys, B. L. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ. Health Perspect.* **123**, 135–43 (2015).
30. Boys, B. *et al.* Fifteen-year global time series of satellite-derived fine particulate matter. *Environ. Sci. Technol.* **1**, 1–27 (2014).

## Author Contributions

D.L.C. conceived the study, performed all analyses, and wrote the manuscript. S.P., A.v.D. and R.V.M. created the exposure models. B.J., S.W., J.B. and B.H. provided advice on the analyses and helped in interpretation of the results. P.P. contributed to the data linkage and preparation of datasets. R.T.B. developed the statistical models and contributed to writing the manuscript. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Crouse, D. L. *et al.* A New Method to Jointly Estimate the Mortality Risk of Long-Term Exposure to Fine Particulate Matter and its Components. *Sci. Rep.* **6**, 18916; doi: 10.1038/srep18916 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>