A time-series analysis of relevant pollutants in Hamilton (Ontario) and induced mortality

The relationship between air pollution and public health has gained increasing attention in the past decade. Many time-series analyses have been conducted worldwide, including in all the major cities of the United States, Europe, and Asia. However, the most current time-series analysis study of Ontario, Canada dates back to 2012 and includes only a single city, calling the need of a more recent study at a provincial scale. As a result, we propose to conduct time-series analyses of major Ontario cities and then use a hierarchical model to pool the results and construct a dose-response relationship and generate a predictive regression.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.1904v1 | CC-BY 4.0 Open Access | rec: 29 Mar 2016, publ: 29 Mar 2016

A time-series analysis of relevant pollutants in Hamilton, ON and induced mortality

Yu Fan Zeng*, Adree Khondker
Bachelor of Health Sciences, Faculty of Health Sciences, McMaster University
(Dated: March 27, 2016)

The relationship between air pollution and public health has gained increasing attention in the past decade. Many time-series analyses have been conducted worldwide, including all the major cities in the United States, Europe, and Asia. However, the most current time-series analysis study of Ontario, Canada dates back to 2012 and is included only a single city, calling the need of a more recent study at a provincial scale. As a result, we propose to conduct time-series analyses of major Ontario cities and then use a hierarchical model to pool the results and construct a dose-response relationship and generate a predictive regression.
Correspondence: zengy10@mcmaster.ca

I. INTRODUCTION

The deleterious effects of high air pollution on public health has been suspected for more than 50 years[1]. In the mid-twentieth century, striking increases in mortality followed a series of severe air pollution incidents in the US and Europe[2, 3]. By the early 1990s, time series studies[4–7] of single locations have demonstrated that even lower air pollution levels increase the rates of mortality and morbidity throughout the developed countries.

Air pollution is due to a heterogeneous mixture of gaseous and particulate components. The main gaseous pollutants are ozone, carbon monoxide, nitrogen oxides, sulfur dioxide, and particulate matter(PM).

Ozone appears naturally in the Earths upper atmosphere and shields organisms on earth from the suns harmful ultraviolet rays[8]. However, near the troposphere and lower atmosphere, ozone irritates the respiratory system, reduces lung function, damages the cells that line the lungs, aggravates asthma and other chronic lung diseases9.

Fine particulate matter, also known as PM 2.5, are particles only detectable using an electron microscope. Major sources of PM2.5 include power plants, wood burning, forest fires, agricultural burning, industrial processes, and motor vehicles[9]. When exposed to particle pollution, patients of heart diseases may experience pain, palpitations, shortness of breath, and fatigue. PM2.5 has been associated with cardiac arrhythmias and heart attacks[10]. Particle pollution also can increase the susceptibility to respiratory infections and can aggravate existing respiratory illnesses for example, asthma and chronic bronchitis[11].

Carbon monoxide is an odorless, colorless gas during incomplete combustion of carbon. According to the United States Environmental Protection Agency, motor vehicle exhaust accounts for 75% of carbon monoxide emissions nationwide[12]. Carbon monoxide enters the bloodstream through the lungs and binds to hemoglobin, the protein molecules that carries oxygen from the lungs to the bodys tissues[13]. Cardiovascular disease patients, such as those with coronary artery disease, are most at risk. They may experience chest pain and other cardiovascular symptoms if they are exposed to high concentrations of carbon monoxide[14].

Sulphur dioxide, a colorless, reactive gas, is emitted when sulphur containing fuels such as coal and oil are burned. Major sources of emissions include refineries, power plants, and industrial boilers. At low concentrations, very brief exposure causes bronchoconstriction in asthma patients accompanied by wheezing, chest tightness, and shortness of breath. Medication is often required to clear the symptoms. At high levels, even healthy individuals will experience similar effects. Long-term exposure to sulfur dioxide may lead to respiratory symptoms and illness, and aggravate asthma.

Nitrogen oxides (NO_x) are emitted as NO but then rapidly reacts with ozone or radicals in the atmosphere to form NO_2. NO_x gases react with precipitation, oxygen, and other atmospheric substances to form smog and acid rain as well as being central to the formation of tropospheric ozone. The major anthropogenic source of nitrogen oxides is the combustion of fossil fuels from the stationary sources (power generation, heating, etc.) and motor vehicles. It can interfere with the bloods ability to carry oxygen through the body, causing headache, fatigue, and dizziness16. Exposure to high concentrations of nitrogen oxides can cause collapse, rapid burning and swelling of tissues in the throat and upper respiratory tract, and fluid build-up in the lungs, and at times even death[15].

There is abundant published work that shows air pollution episodes acutely increases the mortality of exposed populations, as found in [fill in city names and citation]. Theses effects have been related to all five of the main pollutants, but tend to focus on particulate matter, sulfur dioxide, and ozone. On the basis of known effects on community health, an association of exposure with mortality can be expected. However, the pathogenesis of air pollutants diseases is very different from those of acute toxic materials.

Existing epidemiologic studies of air pollution in Ontario and Canadian cities with illnesses are outdated

and incomplete: the most recent study were published in 2012 and most studies consists of a single city or location.

Time-series analysis needs to be updated consistently, and offers a model to extract meaningful forecasts of future trends. Such trends are vital to understand for fields such as public health, occupational health, environmental chemistry, and medicine. Indeed, there is a need for a report for the effects of environmental pollutant levels on mortality for Canada. Dominici et al. have described a model to pool time-series analysis data from different regions to construct a national dose-response relationship between pollutant concentrations and mortality[16]. We will conduct our study following the described method.

II. METHODS

Time series studies aim to associate time-varying pollution exposure with time-varying event counts [17]. These studies assume the health effects are small and disease outcomes are rare, and thus the bias from ignoring data aggregation across individuals should be small[18]. As a sample calculation, we used a generalized linear model (GLM) with natural cubic splines). For sample calculations, please refer to the Sample Calculations section.

The use of multicity studies is a major milestone in time-series research on air pollution. Single-city studies are limited in its usefulness to other areas because the statistical approaches used to analyze raw data vary with each study and characteristics of both the city and its citizens varies dramatically. Hierarchical model in which the estimates are generated by city-specific models can be combined in a second stage to produce a regional or national effect estimate[19].

We will conduct analyze data of all 39 Ontario major cities with data available on the Ontario Ministry of Environment and Climate Change from 2000 2014 and associate pollutant levels with daily mortality using a GLM model with natural cubic splines based on the published method described by Dominici et al[16].

We will first use a two-stage log-linear regression model[20–23]. In the first stage, a separate log-linear x_t regression of the daily mortality rate on the air-pollution measure and other confounders will be fitted to obtain estimates of the relative rate of mortality associated with the pollution variable along with its statistical uncertainty. The outcome variable, $Y_c t^r$, is the total number of mortality on day t, in city c, within region and the exposure variable. C^r, R, and T describes the number of cities within each region r, the number of regions, and the number of days.

$$Y_c t^r | \mu_c t^t = Poisson(\mu_c t^t), \quad (1)$$

$$c = 1, ..., C^r, r = 1, ..., R, t = 1, ...T, \quad (2)$$

$$log \mu_c t^t = {\beta_c}^r PM_{ct-1} + \eta_c X_t \quad (3)$$

$$\mu_c t^r = E[Y_v t^r] \quad (4)$$

X_t is the itth row of the design matrix for the confounding factors (e.g. long-term trends and seasonality in the mortality time series, weather and humidity variables, etc.); $\eta_c$is the corresponding vector of coefficients. The justification for selecting the confounding variables are listed below. These are all potential confounding factors in the calculation of the city-specific relative rates associated with each air pollutant levels[20, 24–27].

During the first-stage of analysis, we will also analyze the effect of the day of pollutant data collection (the current day, the day before, or two days before). We will determine the optimal lag interval to use on the case.

In the second stage, the heterogeneity of the city-specific effects within regions will be described assuming that

$$\beta^r | {\alpha_0}^r, \alpha, \mathbf{Z}^r, \sigma^r = N_{cr}(\alpha_0^r \mathbf{j}^r + \mathbf{Z}^r \alpha, \sigma^2 I) \quad (5)$$

$$r = 1, ..., R, \quad (6)$$

where β^r=[${\beta_1}^r$,. . .,${\beta_(C^r)}^r$] is the collection of true PM10 coefficients for the C^r cities in region r,${\alpha_0}^r$, is the regional air pollution effect when all the covariates are centered at their mean values, j^r is a vector of length C^r having all elements equal to 1, and α=[α_1,. . .,α_p] is the vector of the second-stage regression coefficients (i.e., α_j measures the change in ${\beta_c}^r$ per unit of change in the city-specific covariate Z_{cj}^r), and σ^2 measures the variance of the ${\beta_c}^r$s within each region. The choice of these predictors have been discussed before30. The second-stage covariates are included in the design matrix Z, and the rationale for their inclusion are summarized in the following table.

In the third stage, the estimates of the relative rate will be combined for all cities (after adjustment for the various levels of uncertainty) to yield an overall pooled estimate and to assess whether area-specific characteristics modified the estimated effect of air pollution on the relative rate of death. Assume

$${\alpha_0}^r | \alpha_0, \tau^2 = N(\alpha_0, \tau^2) \quad (7)$$

α_0 is the overall relative rate of mortality for PM10, and τ^2 measures the variance of ${\alpha_0}^r$ across regions.

The vector η_c is multi-dimensional, and a full Bayesian approach to simulate using joint posterior distributions of

${\beta_c}^r$ and η_c and then integrating to obtain the marginal posterior distributions of the ${\beta_c}^r$ is laborious. To simply calculations and save computational times, Dominici and colleagues have identified a method to replace the first stage of the model with the MLE-based normal approximation to the likelihood function4,18.

$$\beta^r | {N_c}^r(\beta^r, V^r) \tag{8}$$

To construct regional dose-response curves of each pollutant,${\beta_c}^r$ PM_{ct-1} in (1) could be replaced with S(PM_{ct-1},knots) in which S is a natural cubic spline with an unknown locations v=(v_1, . . ., v_k) and with boundary knots fixed at specific values for each pollutant.

$$log\mu_c t^t = S(PM_{ct-1}, knots) + \eta_c X_t \tag{9}$$

To primitively understand the change in the number of deaths per pollutant, a relationship between pollutant levels and time (months) was taken and bilinearly interpolated. A three-dimensional plot was then generated with respect to the number of deaths. The data used the regression weight function for data points contained in a span of 25, this prevents the data from being resistant to outliers. Thus, the plot was fitted with a two-dimensional curve and smoothed as a locally weighted scatter plot with MATLAB further described below.

Local regression computes the regression weights, wi, given by the tricube function:

$$w_i = \left(1 - \frac{|x - x_i|}{d(x)}^3\right)^3 \tag{10}$$

Where x is the predictor value, xi are the nearest neighbours of x as defined by the span, and d(x) is the distance along the abscissa from x to the most distant predictor value within the span. A weighted linear least squares regression is done using the first-degree polynomial. The smoothed value is given by the weighted regression.

III. RESULTS AND DISCUSSION

We have conducted a monthly time-series analysis using publicly available data on Ontario Ministry of the Environment and Climate change and Mortality Database, as well as primitive 3D regression with smoothing. All calculations were performed on STATA13 and MATLAB.

IV. SAMPLE TIME SERIES

We used a flexible cubic spline functions with 3 equally spaced reference points. Then the data was fitted in the model and the coefficients of the basis terms are estimated by maximum likelihood such that the linear combination models the seasonal patterns in the outcome data as closely as possible.

The 2011 concentration data for CO, SO_2, and NO_x are missing. As a result, we only conducted analysis from 2003 2010 to study the relationship between mortality and pollutant concentration and 2003-2011 for PM2.5 and ozone. We also adjusted for the effects of temperature using publicly-available historical temperature data.

Our analysis found positive results for four pollutants (CO, SO_2, PM_2.5, and NOx) and one negative result (ozone). The effects of per unit(1 ppm) increase of CO, the chances of mortality increases by almost 20%.

A. Locally Smoothed Regression of Mortality versus Pollutant and Time

Regression was done for three pollutants: CO, SO_2, NO_x. A contour plot is used to show the smoothed regression of the data, whereas the local mean is shown as a 2D surface based on the span. [28] The relationship between the number of deaths and the interpolants of the pollutant and time are taken. From this, there are areas of high local density at time greater than 80, and high relative pollutant concentrations. The contour representing death favors higher time, pollutant concentrations such that the regression suggests there will be increased number of deaths over time and pollutant increases. More analysis needs to be done by fitting Gaussian distributions for a more robust analysis of the trend.

V. ACKNOWLEDGMENTS

We thank Amir Tavassoley, Mark Hwang, Rishi Arora, for valuable discussion and support throughout this project. As well, we thank our local supervisor Laurie Barlow for her critical analysis of this project.

VI. SUPPLEMENTARY INFORMATION

TABLE I. Table caption

Predictors	Reason
Indicator variables for the three age groups	Allow for different baseline mortality rates within each age group
Indicator variables for the day of the week	Allow for different mortality rates within each day of the week
Smooth functions of time with 7 degrees of freedom (df)/year	Adjust for long-term trends and seasonality
Smooth functions of temperature with 6 df	Control for the known effects of temperature
Smooth functions of dewpoint with 3 df	Control for the known effects of humidity
Separate smooth function of time (2 df/yr) for each age group contrast	Separately adjust for seasonality within each age group

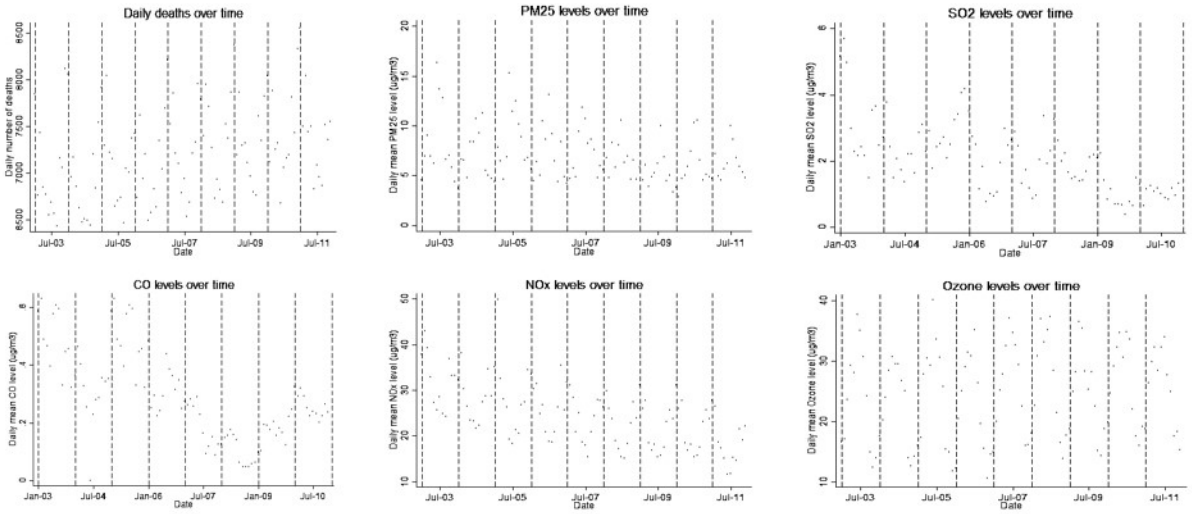


FIG. 1. Plots of mortality over time to pollutant levels over time. Due to seasonal changes, seasonality and long-term trends need to controlled using spline model

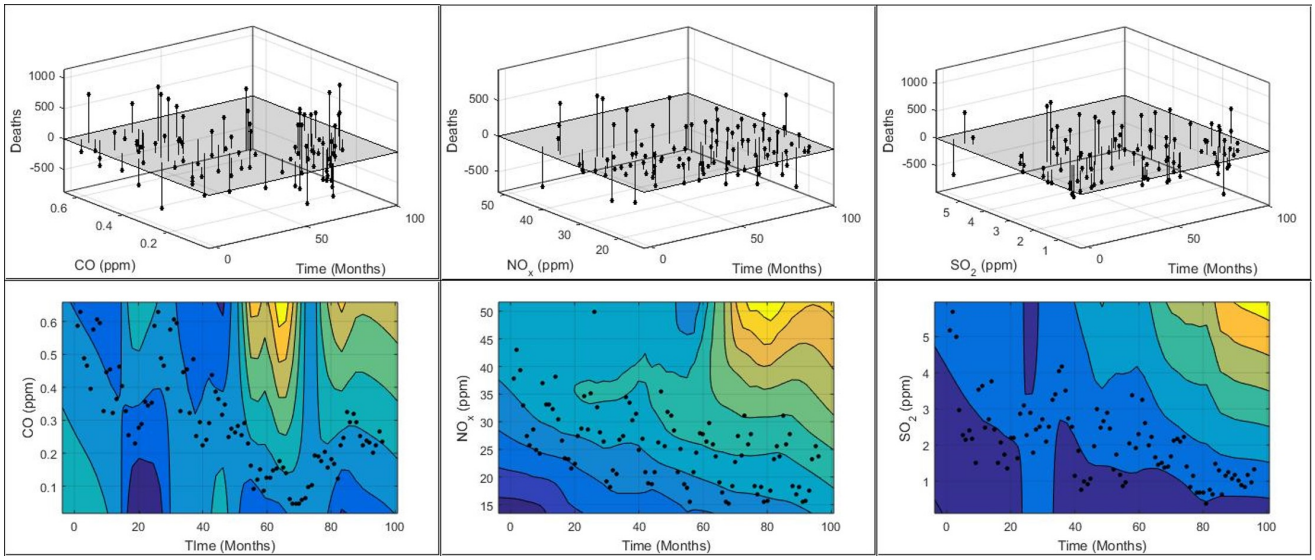
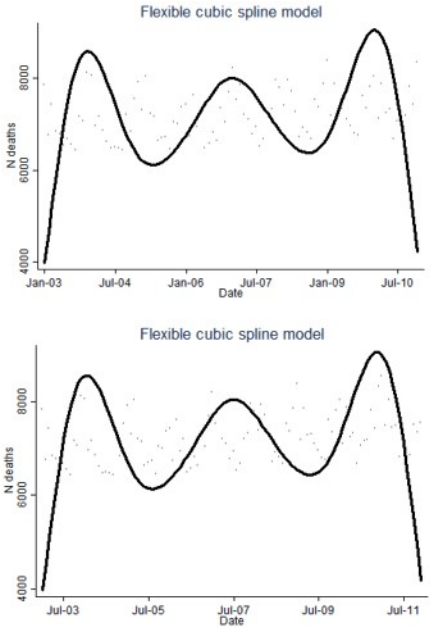


FIG. 2. Contour plot of the smoothed regression of CO, SO_2, and NOx data. The local means is shown as 2D surface based on the psan. The regression suggests there will be increased number of deaths over time and pollutant increases. Next step of the analysis is fitting Gaussian distribution for a more robust analysis of the trend.



Death	IRR	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
NOxdivided10	1.109642	.0514698	2.24	0.025	1.013213	1.215248

Death	IRR	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
CO	1.197534	.2405282	0.90	0.369	.807833	1.775228

Death	IRR	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
SO2	1.10035	.0266744	3.94	0.000	1.049292	1.153893

Death	IRR	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Ozonedivided10	.9855139	.026769	-0.54	0.591	.9344198	1.039402

Death	IRR	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
PM25divided10	1.123523	.1166676	1.12	0.262	.9166264	1.377119

FIG. 3. Cubic spline model with three knots for both studies from 2003 - 2010 and 2003 - 2011. The tables shows the relatiave increase risk of mortality per unit increase of each pollutant, along with their 95% confidence interval. Our analysis found positive results for four pollutants (CO, SO_2, $PM_{2.5}$, and NOx) and one negative result (ozone). An increase in 10 parts per billion (ppb) of nitric oxides will leads to a 11% increase in the risk of mortality. 1 ppm increase is associated with 19.8% increase of mortailty relative risks. For sulpfur dioxide, a unit increase (1 ppb) is associated with 10% increase in the rate or mortality, while 10 ppb increase of ozone is associated with a 2% decrease in mortality risks. A 10 unit(mg/cubic cm) increase in PM 2.5 levels is associated with a 12% increase in the risks of mortality. However, due to that only monthly mortality data were available to us, the sample size is small and as a result, the study has fairly large confidence intervals. Another limitation is that we lacked access to socioeocnomic data of the population to fruther control for the influences of confounding variables.

[1] B. Chardon, S. Host, G. Pedrono, and I. Gremy, "Apport de la méthode cas-croisé à lanalyse des effets sanitaires à court terme de la pollution atmosphérique: réanalyse de données du programme erpurs," *Revue d'Épidémiologie et de Santé Publique*, vol. 56, no. 1, pp. 31–40, 2008.

[2] A. Ciocco and D. J. Thompson, "A follow-up of donora ten years after: methodology and findings," *American Journal of Public Health and the Nations Health*, vol. 51, no. 2, pp. 155–164, 1961.

[3] W. Logan, "Mortality in the london fog incident, 1952," *The Lancet*, vol. 261, no. 6755, pp. 336–338, 1953.

[4] F. W. Lipfert and R. E. Wyzga, "Air pollution and mortality: issues and uncertainties," *Journal of the Air & Waste Management Association*, vol. 45, no. 12, pp. 949–966, 1995.

[5] C. A. Pope III, M. J. Thun, M. M. Namboodiri, D. W. Dockery, J. S. Evans, F. E. Speizer, and C. W. Heath Jr, "Particulate air pollution as a predictor of mortality in a prospective study of us adults," *American journal of respiratory and critical care medicine*, vol. 151, no. 3_pt_1, pp. 669–674, 1995.

[6] J. M. Samet, M. C. Marbury, and J. D. Spengler, "Health effects and sources of indoor air pollution. part i," *American Review of Respiratory Disease*, vol. 136, no. 6, pp. 1486–1508, 1987.

[7] R. M. Doherty, "Atmospheric chemistry: Ozone pollution from near and far," *Nature Geoscience*, vol. 8, no. 9, pp. 664–665, 2015.

[8] M. Lippmann, "Health effects of tropospheric ozone: review of recent research findings and their implications to ambient air quality standards.," *Journal of exposure analysis and environmental epidemiology*, vol. 3, no. 1, pp. 103–129, 1992.

[9] B. Ostro, L. Chestnut, N. Vichit-Vadakan, and A. Laixuthai, "The impact of particulate matter on daily mortality in bangkok, thailand," *Journal of the Air & Waste Management Association*, vol. 49, no. 9, pp. 100–107, 1999.

[10] N. Vichit-Vadakan, B. D. Ostro, L. G. Chestnut, D. M. Mills, W. Aekplakorn, S. Wangwongwatana, and N. Panich, "Air pollution and respiratory symptoms: results from three panel studies in bangkok, thailand.," *Environmental Health Perspectives*, vol. 109, no. Suppl 3, p. 381, 2001.

[11] U. Govt, "United states enviromental protection agency. cars, trucks, buses, and nonroad equipment,"

[12] L. K. Weaver, R. O. Hopkins, K. J. Chan, S. Churchill, C. G. Elliott, T. P. Clemmer, J. F. Orme Jr, F. O. Thomas, and A. H. Morris, "Hyperbaric oxygen for acute carbon monoxide poisoning," *New England Journal of Medicine*, vol. 347, no. 14, pp. 1057–1067, 2002.

[13] E. N. Allred, E. R. Bleecker, B. R. Chaitman, T. E. Dahms, S. O. Gottlieb, J. D. Hackney, M. Pagano, R. H. Selvester, S. M. Walden, and J. Warren, "Short-term effects of carbon monoxide exposure on the exercise performance of subjects with coronary artery disease," *New England journal of medicine*, vol. 321, no. 21, pp. 1426–1432, 1989.

[14] A. L. Pruden and D. F. Ollis, "Photoassisted heterogeneous catalysis: the degradation of trichloroethylene in water," *Journal of Catalysis*, vol. 82, no. 2, pp. 404–417, 1983.

[15] F. Dominici, M. Daniels, S. L. Zeger, and J. M. Samet, "Air pollution and mortality: estimating regional and national dose-response relationships," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 100–111, 2002.

[16] M. L. Bell, J. M. Samet, and F. Dominici, "Time-series studies of particulate matter," *Annu. Rev. Public Health*, vol. 25, pp. 247–280, 2004.

[17] J. Wakefield and R. Salway, "A statistical framework for ecological and aggregate studies," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 164, no. 1, pp. 119–137, 2001.

[18] E. Samoli, A. Analitis, G. Touloumi, J. Schwartz, H. R. Anderson, J. Sunyer, L. Bisanti, D. Zmirou, J. M. Vonk, J. Pekkanen, *et al.*, "Estimating the exposure-response relationships between particulate matter and mortality within the aphea multicity project," *Environmental Health Perspectives*, pp. 88–95, 2005.

[19] J. M. Samet, F. Dominici, F. C. Curriero, I. Coursac, and S. L. Zeger, "Fine particulate air pollution and mortality in 20 us cities, 1987–1994," *New England journal of medicine*, vol. 343, no. 24, pp. 1742–1749, 2000.

[20] T. J. Hastie and R. J. Tibshirani, *Generalized additive models*, vol. 43. CRC Press, 1990.

[21] D. V. Lindley and A. F. Smith, "Bayes estimates for the linear model," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–41, 1972.

[22] M. J. Fazzari and J. M. Greally, "Introduction to epigenomics and epigenome-wide analysis," in *Statistical Methods in Molecular Biology*, pp. 243–265, Springer, 2010.

[23] I. Mavrin, D. Knežević, and I. Jurić, "Impact of diesel engine exhaust gases on environmental pollution and human health," *PROMET-Traffic&Transportation*, vol. 16, no. 4, pp. 197–205, 2012.

[24] J. Kelsall, J. Samet, S. Zeger, and J. Xu, "Air pollution and mortality in philadelphia, 1974–1988," *American journal of epidemiology*, vol. 146, no. 9, pp. 750–762, 1997.

[25] J. Samet, S. Zeger, J. Kelsall, J. Xu, and L. Kalkstein, "Does weather confound or modify the association of particulate air pollution with mortality?: an analysis of the philadelphia data, 1973–1980," *Environmental Research*, vol. 77, no. 1, pp. 9–19, 1998.

[26] F. Dominici, J. M. Samet, and S. L. Zeger, "Combining evidence on air pollution and daily mortality from the 20 largest us cities: a hierarchical modelling strategy," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 163, no. 3, pp. 263–302, 2000.

[27] J. Samet, F. Dominici, S. Zeger, J. Schwartz, and D. Dockery, "The national morbidity, mortality, and air pollution study. part i: Methods and methodologic issues.," *Research report (Health Effects Institute)*, no. 94 Pt 1, pp. 5–14, 2000.

[28] S. M. Edwards, Z. Kote-Jarai, J. Meitz, R. Hamoudi, Q. Hope, P. Osin, R. Jackson, C. Southgate, R. Singh, A. Falconer, *et al.*, "Two percent of men with early-onset prostate cancer harbor germline mutations in the brca2 gene," *The American Journal of Human Genetics*, vol. 72, no. 1, pp. 1–12, 2003.