

RESEARCH

Open Access



A tensor product quasi-Poisson model for estimating health effects of multiple ambient pollutants on mortality

Li-Jun Xu[†], Shuang-Quan Shen[†], Li Li, Ting-Ting Chen, Zhi-Ying Zhan and Chun-Quan Ou^{* ID}

Abstract

Background: People are exposed to mixtures of highly correlated gaseous, liquid and solid pollutants. However, in previous studies, the assessment of air pollution effects was mainly based on single-pollutant models or was simultaneously included as multiple pollutants in a model. It is essential to develop appropriate methods to accurately estimate the health effects of multiple pollutants in the presence of a high correlation between pollutants.

Methods: The flexible tensor product smooths of multiple pollutants was applied for the first time in a quasi-Poisson model to estimate the health effects of SO₂, NO₂ and PM₁₀ on daily all-cause deaths during 2005–2012 in Guangzhou, China. The results were compared with those from three other conventional models, including the single-pollutant model and the three-pollutant model with and without first-order interactions.

Results: The tensor product model revealed a complex interaction among three pollutants and significant combined effects of PM₁₀, NO₂ and SO₂, which revealed a 2.53% (95%CI: 1.03–4.01%) increase in mortality associated with an interquartile-range (IQR) increase in the concentrations of all three pollutants. The combined effect estimated by the single-pollutant model was 5.63% (95% CI: 3.96–7.34%). Although the conventional three-pollutant models produced combined effect estimates (2.20, 95%CI, 1.18–3.23%; 2.78, 95%CI: 1.35–4.23%) similar to those of the tensor product model, they distorted the estimates and inflated the variances of the estimates when attributing the combined health effects to individual pollutants.

Conclusions: The single-pollutant model or conventional multi-pollutant model may yield misleading results in the presence of collinearity. The tensor product quasi-Poisson regression provides a novel approach to the assessment of the health impacts of multiple pollutants by flexibly fitting the interaction effects and avoiding the collinearity problem.

Keywords: Tensor product, Quasi-Poisson regression, Collinearity, Health effects

Background

An increasing number of studies have demonstrated the association between daily ambient air pollution levels and adverse health outcomes worldwide [1–4]. Some studies generally estimated the health effect of each pollutant separately via a single-pollutant model controlling for time trends and meteorological factors [5–7]. However, pollution occurs as a mixture of highly correlated gaseous, liquid and solid pollutants. Some studies observed that the effect using the single-pollutant model was weakened after

other pollutants were simultaneously included [8]. Therefore, the results of the single-pollutant model tends to overestimate the health effects of pollutants. Moreover, many studies observed synergistic effects of combinations of pollutants on health outcomes [9, 10]. To advance our understanding of the biological mechanisms of pollutant toxicity, it is essential to move from the single-pollutant model to the multi-pollutant model [11]. This would provide important information for guiding regulatory policies for public health.

Many researchers have recognized the limitations of the single-pollutant model and have taken some efforts to overcome such limitations. One of commonly used methods is to simultaneously include two or three

* Correspondence: ouchunquan@hotmail.com

[†]Li-Jun Xu and Shuang-Quan Shen contributed equally to this work. State Key Laboratory of Organ Failure Research, Department of Biostatistics, Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou 510515, China



pollutant variables in a generalized additive model (GAM); then, the effects of one pollutant can be estimated after adjusting for other pollutants in the model [12–14]. This simple resolution ignores the high correlation of the occurrence of pollutants, which likely leads to a collinearity problem, such as variance inflation or bias even reverse estimation of parameters. For example, Chen et al. [15] and Kan et al. [16] showed that PM₁₀ was negatively associated with daily respiratory mortality when the NO₂ or SO₂ was considered simultaneously, suggesting that the PM₁₀ had a protective effect. To alleviate this problem, the Health Effects Institute supports the development of innovative statistical methods for studying the health effects of multiple pollutants in the presence of collinearity.

Several different approaches have been attempted to analyze the effects of multiple pollutants. One approach is the principal component analysis (PCA) that can avoid unstable parameter estimates in the condition of collinearity [17, 18]. However, PCA only maximizes the variance interpreted by the linear combination of independent variables without directly considering the relationship between the dependent variable and independent variables. Partial least-squares regression (PLSR) combining PCA with canonical correlation analysis captures not only the information of independent variables but also the relationship with the dependent variable, yet this approach produces linear combinations of the original independent variables that are difficult to interpret in practice. Using Monte Carlo simulations, Månsson and Shukur [19] showed that the estimators produced by Poisson ridge regression were more accurate (i.e. had less standard errors) than those produced by the Poisson log-linear model. However, the increase in accuracy was at the expense of increasing bias in the condition of strong collinearity. Bayesian model averaging (BMA) averages the effects of models for all possible combinations of independent variables; these models are weighted by the posterior probability of the model [20]. However, there are different views on the interpretation of effect estimates from BMA in the case of extensive collinearity. Bobb et al. [21] proposed Bayesian kernel machine regression (BKMR) that regressed the health outcome on a flexible kernel function of multiple pollutants. BKMR can be highly sensitive to the specification of the prior distribution and can solve the problem of collinearity using hierarchical variable selection.

To date, there is no standard model for evaluating health effects of multiple pollutants. Sun et al. [22] concluded that no method was unanimous across all simulation scenarios that differed in sample size, the number of pollutants and the strength of exposure-response association. The choice of the appropriate method should depend on the goal of the study. Wood

proposed a general method for constructing low-rank tensor product smooths of multiple variables that could quantitatively estimate the nonlinear effects of multiple independent variables on a dependent variable [23, 24]. The tensor product smooths has been widely applied in geometry, physics, mechanics and quantum theory [25]. To the best of our knowledge, the tensor product smooths for GAM was never employed to evaluate the health impacts of multiple pollutants. This study employed firstly this approach to assess non-linear effects and potential interaction effects of SO₂, NO₂ and PM₁₀ on mortality in Guangzhou, China, during 2005–2012.

Methods

Data

Daily meteorological data for mean temperature, relative humidity and atmospheric pressure were collected at the Guangzhou Weather Station and were downloaded from the China Meteorological Data Sharing Service System. Air pollution data, including particulate matter < 10 μm in aerodynamic diameter (PM₁₀), nitrogen dioxide (NO₂) and sulfur dioxide (SO₂), were provided by the Guangzhou Bureau of Environmental Protection. The daily concentrations of PM₁₀, NO₂ and SO₂ were the average levels collected from seven fixed-site air quality monitoring stations that were all located in central urban areas and had complete data during the study period of 2005–2012 [26]. There were 40 missing observations for PM₁₀, NO₂, and SO₂, accounting for 1.5% of the study days. We did not replace the missing values. Because the levels of air pollution and health effects of pollution on mortality may be different in suburban areas, to maintain consistency of the study population, the mortality data were also limited to death registration records of the whole population at six central urban districts.

Statistical analysis

First, we constructed a basic quasi-Poisson regression model without air pollutant variables. Natural cubic splines (ns) of daily mean relative humidity, atmospheric pressure and temperature were used to control the nonlinear and non-monotonic confounding effects of meteorological factors on mortality. Time can be considered as a proxy of some other unobserved time-varying confounders. A natural cubic spline function of time was used to control for the long-term trend and seasonal variations of daily mortality and potential confounding effects of unobserved time-varying covariates. The partial autocorrelation function (PACF) of residuals and the Akaike's Information Criterion (AIC) were used to guide the selection of the number degrees of freedom (df) for ns. When a df of 7 was used for time variables, the PACF was free of patterns and was not autocorrelated [16, 27]. The value of df equal to 3 was used for meteorological variables because this value

has been shown to control well their impacts on daily mortality changes [28–30]. We used a 14-day moving average of mean temperature to control the lag effect because previous studies in Guangzhou observed that the effects of mean temperature on mortality persisted for approximately 14 days [31, 32]. Day of the week (DOW) and a holiday (HOD) indicator were included in the model as indicator variables.

After the basic model was constructed, we added air pollution variable(s) to the model. The reason was that many studies, including our data, showed that the adverse impacts of each specific pollutant were statistically significant at lag 0 and lag 1, and these impacts almost disappeared at lag 3 and during longer lags [33, 34]. To adequately capture the lag effects, a 3-day moving average of air pollution concentrations (i.e., lag 0–2) was used, and the concentrations were introduced into the model in the following four ways:

Model I (single-pollutant model) incorporated each of pollutant variables into the basic model, separately:

$$\begin{aligned} \text{Log}(\mu_t) = & \alpha + \theta_1 \text{ns}(\text{Temperature}_t) + \theta_2 \text{ns}(\text{Humidity}_t) \\ & + \theta_3 \text{ns}(\text{Pressure}_t) + \theta_4 \text{ns}(\text{Time}) + \delta_1 \text{DOW}_t \\ & + \delta_2 \text{HOD}_t + \beta(\text{PM}_{10t}/\text{NO}_{2t}/\text{SO}_{2t}), \end{aligned} \tag{1}$$

where $\beta(\text{PM}_{10t}/\text{NO}_{2t}/\text{SO}_{2t})$ signifies that respective single-pollutant models for PM_{10} , NO_2 and SO_2 are established; β is the coefficient of a pollutant variable.

Model II (three-pollutant non-interaction model), simultaneously included three pollutant variables without considering first-order interactions and is specified as follows:

$$\begin{aligned} \text{Log}(\mu_t) = & \alpha + \theta_1 \text{ns}(\text{Temperature}_t) + \theta_2 \text{ns}(\text{Humidity}_t) \\ & + \theta_3 \text{ns}(\text{Pressure}_t) + \theta_4 \text{ns}(\text{Time}) + \delta_1 \text{DOW}_t \\ & + \delta_2 \text{HOD}_t + \beta_1 \text{PM}_{10t} + \beta_2 \text{NO}_{2t} + \beta_3 \text{SO}_{2t}. \end{aligned} \tag{2}$$

Model III (three-pollutant interaction model) adds linear terms of the first-order interaction between pollutants to model II:

$$\begin{aligned} \text{Log}(\mu_t) = & \alpha + \theta_1 \text{ns}(\text{Temperature}_t) + \theta_2 \text{ns}(\text{Humidity}_t) \\ & + \theta_3 \text{ns}(\text{Pressure}_t) + \theta_4 \text{ns}(\text{Time}) + \delta_1 \text{DOW}_t \\ & + \delta_2 \text{HOD}_t + \beta_1 \text{PM}_{10t} + \beta_2 \text{NO}_{2t} + \beta_3 \text{SO}_{2t} \\ & + \beta_{12}(\text{PM}_{10t} \times \text{NO}_{2t}) + \beta_{13}(\text{PM}_{10t} \times \text{SO}_{2t}) \\ & + \beta_{23}(\text{NO}_{2t} \times \text{SO}_{2t}), \end{aligned} \tag{3}$$

where β_1, β_2 and β_3 are the regression coefficients of pollutants, and β_{12}, β_{13} and β_{23} are the regression coefficients of first-order interaction terms.

Model IV (tensor product model) constructs the tensor product smooths (te) of three pollutants (PM_{10} , NO_2 and SO_2) and is specified as follows:

$$\begin{aligned} \text{Log}(\mu_t) = & \alpha + \theta_1 \text{ns}(\text{Temperature}_t) + \theta_2 \text{ns}(\text{Humidity}_t) \\ & + \theta_3 \text{ns}(\text{Pressure}_t) + \theta_4 \text{ns}(\text{Time}) + \delta_1 \text{DOW}_t \\ & + \delta_2 \text{HOD}_t + te(\text{PM}_{10}, \text{NO}_2, \text{SO}_2), \end{aligned} \tag{4}$$

where $te(\text{PM}_{10}, \text{NO}_2, \text{SO}_2)$ is the tensor product smooth function of three pollutants. For models I-IV, μ_t is the expected number of deaths on day t , α is the intercept, $\text{ns}(\cdot)$ denotes the unknown smooth functions modeled by natural cubic splines, and θ_1 - θ_4 represent the coefficients vector of smooth functions.

Construction of the tensor product smooths has been described previously in detail [23]. Mathematically, supposing that there are two vector spaces \mathbf{V} and \mathbf{W} , where the \mathbf{V} has a basis e_i ($i = 1 \dots m$), and the \mathbf{W} has a basis f_j ($j = 1 \dots n$), the tensor product of $\mathbf{V} \otimes \mathbf{W}$ is generated by $\sum_{i=1}^m \sum_{j=1}^n e_i \otimes f_j$, where the product operation \otimes is Kronecker product. In this paper, we briefly introduce the construction process of tensor product smooths of three covariates (i.e., PM_{10} , NO_2 , SO_2). It is easy to generalize the tensor product smooths to more variables in the same way. Basis functions for representing the smooth function of each pollutant are given by

$$f(\text{PM}_{10}) = \sum_{i=1}^I \alpha_i \phi_i(\text{PM}_{10}), \quad f(\text{NO}_2) = \sum_{k=1}^K \omega_k \eta_k(\text{NO}_2)$$

and $f(\text{SO}_2) = \sum_{l=1}^L b_l \phi_l(\text{SO}_2)$, where the $\phi_i(\text{PM}_{10})$, $\phi_l(\text{SO}_2)$ and $\eta_k(\text{NO}_2)$ are the basis functions and α_i , b_l and ω_k are parameters. Next, the smooth of PM_{10} , denoted by $f(\text{PM}_{10})$, is converted into a smooth function of PM_{10} and SO_2 , denoted by $f(\text{PM}_{10}, \text{SO}_2)$ (to make $f(\text{PM}_{10})$ change smoothly with SO_2), which can be achieved by allowing α_i to vary smoothly with SO_2 . Thus, we obtain

$$\alpha_i(\text{SO}_2) = \sum_{l=1}^L b_{il} \phi_l(\text{SO}_2)$$

which results in

$$\begin{aligned} f(\text{PM}_{10}, \text{SO}_2) = & \sum_{i=1}^I \alpha_i(\text{SO}_2) \phi_i(\text{PM}_{10}) \\ = & \sum_{i=1}^I \sum_{l=1}^L b_{il} \phi_l(\text{SO}_2) \phi_i(\text{PM}_{10}). \end{aligned}$$

Similarly, we can now create a smooth function of PM_{10} , SO_2 and NO_2 by allowing $f(\text{PM}_{10}, \text{SO}_2)$ to vary smoothly with NO_2 . Therefore, with following the same reasoning as before, we obtain

$$te(\text{PM}_{10}, \text{NO}_2, \text{SO}_2) = f(\text{PM}_{10}, \text{NO}_2, \text{SO}_2) = \sum_{i=1}^I \sum_{l=1}^L \sum_{k=1}^K v_{ilk} \eta_k(\text{NO}_2) \phi_l(\text{SO}_2) \phi_i(\text{PM}_{10}),$$

where v_{ilk} are the coefficients of basis functions of tensor product smooths.

According to the above description, it is clear that the complexity of the tensor product smooth function depends on the type of basis function and its dimensionality; therefore, it is essential to choose an appropriate basis function describing the relationship between an adverse outcome and air pollutants. In this study, we used the cubic spline basis function that is widely used to flexibly fit a potential nonlinear relationship.

To avoid the problem of collinearity and controlling the tradeoff between goodness of fit and model smoothness, we define the following penalized likelihood function to estimate parameters:

$$l(v) + J(f(\text{PM}_{10}, \text{NO}_2, \text{SO}_2))$$

(and)

$$J(f(\text{PM}_{10}, \text{NO}_2, \text{SO}_2)) = \int_{\text{PM}_{10}, \text{SO}_2, \text{NO}_2} [\lambda_{\text{PM}_{10}} \left(\frac{\partial^2 f}{\partial (\text{PM}_{10})^2} \right)^2 + \lambda_{\text{NO}_2} \left(\frac{\partial^2 f}{\partial (\text{NO}_2)^2} \right)^2 + \lambda_{\text{SO}_2} \left(\frac{\partial^2 f}{\partial (\text{SO}_2)^2} \right)^2] d\text{PM}_{10} d\text{NO}_2 d\text{SO}_2,$$

where $l(v)$ is the likelihood function, and $J(f(\text{PM}_{10}, \text{NO}_2, \text{SO}_2))$ represents tensor product penalties. $\lambda_{\text{PM}_{10}}$, λ_{NO_2} and λ_{SO_2} are smoothing parameters. Such parameters can be selected by Generalized Cross Validation (GCV). The estimates, denoted by \hat{v}_{ilk} , are obtained by penalized iteratively re-weighted least squares which maximizes the penalized likelihood function. Furthermore, stratified analyses by educational attainment, sex and age were conducted respectively.

Estimation of the combined effects

The combined effects of three air pollutants were measured by the combined rate ratio (RR). For the single-pollutant model and the three-pollutant non-interaction model, RR was calculated as the exponentiated sum of the products of each pollutant's regression coefficient and the pollutant's increment. For the three-pollutant interaction model, RR was calculated as the exponentiated sum of the following two parts: 1) the product of each pollutant's regression coefficient and the pollutant's increment and 2) the product of each interaction term coefficient and the difference value of the interaction term. The standard error of RR was estimated based on the variance-covariance matrix of pollutants' coefficients. The detailed derivation process is described in the supplemental materials and uses the method proposed

by Winquist et al.[13]. For the tensor production model, the RR is calculated as follows:

$$te^T v = te(\text{PM}_{10}^{j^{\text{th}}}, \text{NO}_2^{j^{\text{th}}}, \text{SO}_2^{j^{\text{th}}}) - te(\text{PM}_{10}^{i^{\text{th}}}, \text{NO}_2^{i^{\text{th}}}, \text{SO}_2^{i^{\text{th}}}),$$

where

$$te(\text{PM}_{10}^{j^{\text{th}}}, \text{NO}_2^{j^{\text{th}}}, \text{SO}_2^{j^{\text{th}}}) = \sum_{i=1}^I \sum_{l=1}^L \sum_{k=1}^K v_{ilk} \eta_k(\text{NO}_2^{j^{\text{th}}}) \phi_l(\text{SO}_2^{j^{\text{th}}}) \phi_i(\text{PM}_{10}^{j^{\text{th}}});$$

$$te(\text{PM}_{10}^{i^{\text{th}}}, \text{NO}_2^{i^{\text{th}}}, \text{SO}_2^{i^{\text{th}}}) = \sum_{i=1}^I \sum_{l=1}^L \sum_{k=1}^K v_{ilk} \eta_k(\text{NO}_2^{i^{\text{th}}}) \phi_l(\text{SO}_2^{i^{\text{th}}}) \phi_i(\text{PM}_{10}^{i^{\text{th}}}).$$

Therefore, we can get $RR = \exp(te^T v)$.

The values $te(\text{PM}_{10}^{j^{\text{th}}}, \text{NO}_2^{j^{\text{th}}}, \text{SO}_2^{j^{\text{th}}})$ and $te(\text{PM}_{10}^{i^{\text{th}}}, \text{NO}_2^{i^{\text{th}}}, \text{SO}_2^{i^{\text{th}}})$ are the values of the tensor product smooths at the j^{th} and i^{th} percentiles of pollutants' concentrations, respectively. The vector v is the coefficient vector of the basis function of tensor product smooths. The vector te^T is the transposed difference vector of basis functions of tensor product smooths at the j^{th} percentiles and the i^{th} percentiles of pollutants' concentrations. The 95% confidence interval (CI) of RR was estimated by the following formula:

$$95\%CI = (\exp(te^T v - z_{1-\alpha/2} se_{te^T v}), \exp(te^T v + z_{1-\alpha/2} se_{te^T v})),$$

$$se_{te^T v} = \sqrt{te^T \hat{\Sigma} te},$$

where $\hat{\Sigma}$ is the variance-covariance of the coefficients of the basis function and $\alpha = 0.05$.

To improve the readability of the results and better understand the excess burden due to air pollution, we calculated the excess risk (ER) of mortality by $(RR-1)*100\%$, i.e., the percentage increase in mortality associated with air pollution.

To clarify the robustness of the lag selection for air pollutants, we conducted a sensitivity analysis to determine the effects of air pollution at lag 0–1 by using a 2-day moving average in models I-IV.

All analyses were performed using R software version 3.3.1, by using the “mgcv” used to fit the GAM.

Results

The total number of all-cause deaths was 193,715 during 2005–2012, with 66.34 cases per day. The daily mean temperature had an average of 22.5 °C and ranged from 5.1 °C to 34.2 °C. The mean daily average concentrations of PM₁₀, SO₂ and NO₂ were 74.9 µg/m³, 41.0 µg/m³ and 62.4 µg/m³, respectively. The interquartile ranges (IQR) of three pollutants were 48.2 µg/m³, 33.7 µg/m³ and 33.3 µg/m³, respectively (Table 1). The time series plot

Table 1 Summary statistics for daily number of deaths, daily air pollution concentrations and weather conditions in Guangzhou, China, 2005–2012

Variables	Mean ± SD	Min	Percentile			Max
			25th	50th	75th	
Daily number of deaths	66 ± 14	21	56	64	75	248
PM ₁₀ (µg/m ³)	74.9 ± 39.6	7.6	46.4	66.9	94.6	342.3
SO ₂ (µg/m ³)	41.0 ± 29.6	2.3	19.9	33.7	53.6	214.1
NO ₂ (µg/m ³)	62.4 ± 27.8	16.7	42.1	55.6	75.4	254.7
Mean temperature (°C)	22.5 ± 6.3	5.1	18.0	24.2	27.7	34.2
Mean humidity (%)	72.9 ± 13.2	20.0	65.0	75.0	83.0	99.0
Mean pressure (hpa)	10,074 ± 70	9874	10,023	10,071	10,128	10,272

of air ambient pollutants and all-cause deaths revealed a consistent seasonality; there were generally more deaths and higher concentrations of pollutants in the cool season (from May to October) than in the warm season (from November to April) (Fig. 1). The Spearman correlations of PM₁₀ and NO₂, PM₁₀ and SO₂, and NO₂ and SO₂ were 0.83, 0.60 and 0.65, respectively ($P < 0.01$).

The tensor product model can produce a four-dimensional plot for the RR of mortality associated with the levels of three pollutants. To intuitively show the combined effects of air pollutants, the concentration of one of pollutants was fixed at the reference value (the 25th percentile was used in this study); then, a three-dimensional graph of RR of the other two pollutants can be drawn (Fig. 2). Generally, as the concentration of pollutants increased, a stronger combined effect of pollutants on mortality was observed, but the relationship was not completely linear or monotonic. The effect of one pollutant varied with the concentration of another pollutant. For example, the effect of PM₁₀ increased significantly with the concentration of SO₂ at low concentrations, but the curve of PM₁₀ effect was relatively flat for high concentrations of SO₂, which indicated that there was an interaction between pollutants.

Table 2 and Fig. 3 show the effects of air pollution estimated by four different models. The single-pollutant model (Model I) revealed that an IQR increase in the concentration of PM₁₀, NO₂ and SO₂ was associated with increases in mortality of 2.03% (95% CI: 1.08–2.99%), 1.63% (95% CI: 0.69–2.59%) and 1.86% (95% CI: 0.94–2.78%), respectively. The level of NO₂ was negatively associated with the risk of mortality if PM₁₀ and SO₂ were included simultaneously in the non-interaction model (models II–III). The combined effect of three pollutants estimated by the single-pollutant model (ER = 5.63, 95%CI: 3.96–7.34%) was much greater than those obtained from other three multi-pollutant models. The combined effect estimated by the three-pollutant non-interaction model (ER = 2.20, 95%CI: 1.18–3.23%) was slightly smaller than those from the interaction model (ER = 2.78, 95%CI: 1.35–4.23%) and

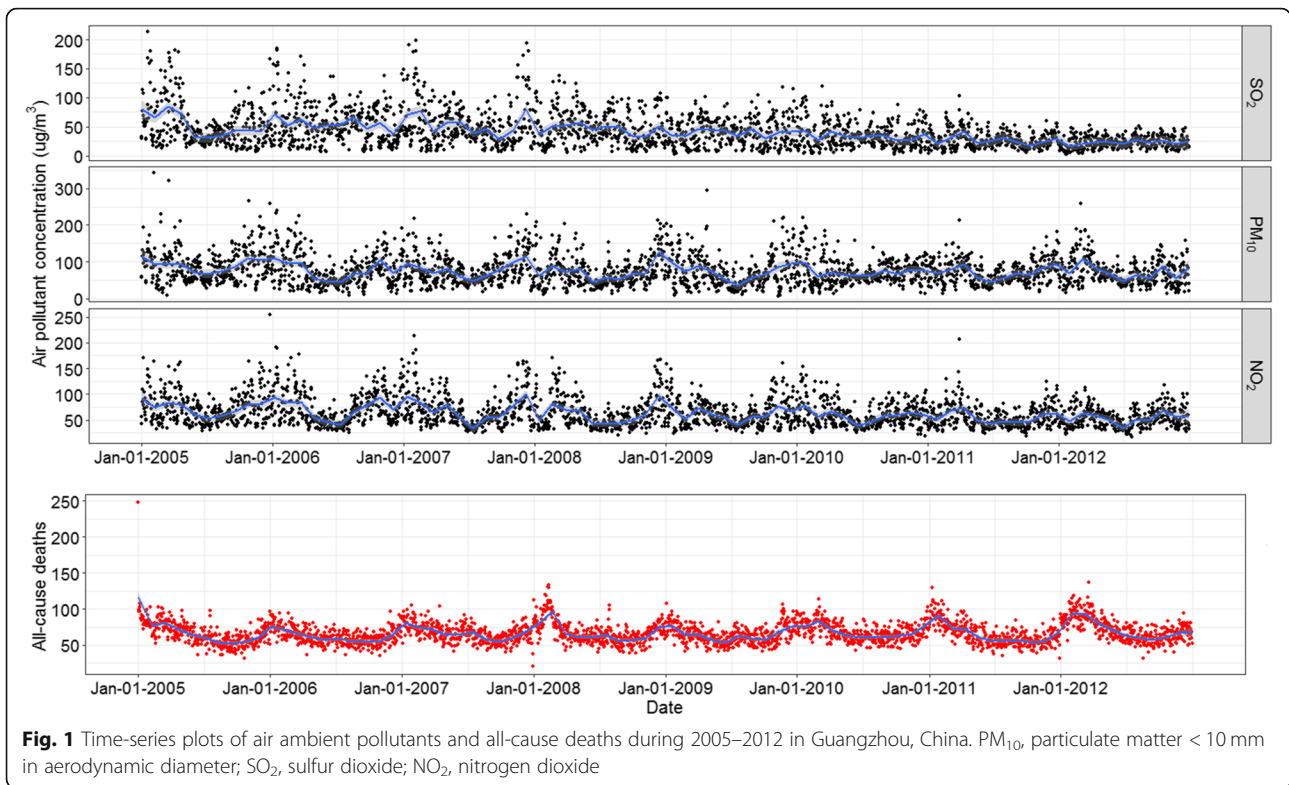
the tensor product model (ER = 2.53, 95%CI: 1.03–4.01%). Fig. 3(a–c) shows that the tensor product model, compared to the three-pollutant non-interaction model and interaction model, produced more precise effect estimates (i.e., narrower 95%CI), especially for the main effects of single pollutants.

Table 3 shows the results of stratified analyses by the tensor product model. There were a total number of 193,715 deaths during 2005–2012 in Guangzhou; 24.4, 56.4 and 92.6% of deaths were of those aged under 65, males and individuals with low educational attainment (illiterate or only primary education), respectively. Females, elderly people (65 years of age or above) and individuals with low educational attainment were more vulnerable to ambient air pollution exposure than were males, younger individuals and those with high educational attainment.

The sensitivity analysis for all models revealed that the effects at lag 0–1 were slightly larger than those at lag 0–2 (Additional file 1: Figure S1).

Discussion

Most of previous studies of the health effects of air pollution are based on a single-pollutant model or include multiple pollutants in a model at the same time, thus ignoring the possibility of high correlation among pollutants that leads to the inflation of variance and the instability of model parameter estimation. The commonly used solution is stepwise regression and removing one of the highly correlated variables; however, this method may cause the loss of some important information. In this study, we proposed a quasi-Poisson regression with tensor product smooth to estimate the health effects of multiple pollutants. Overall, we observed that the combined effect of three pollutants (PM₁₀, SO₂ and NO₂) gradually increased along with concentration of pollutants. Moreover, there was a complex interaction among the pollutants. The results of the stratified analysis showed that elderly individuals, females and people with low educational attainment were more vulnerable to air pollution.



The single-pollutant model, a common method of assessing health effects of air pollution, revealed that the levels of each single pollutant (PM₁₀, NO₂ and SO₂) had a significantly positive relationship with the risk of all-cause deaths, with ER values associated with IQR increase in the concentration of lag 0–2 of 2.03, 1.63 and 1.86%, respectively. We observed that the effect estimates of PM₁₀ and NO₂ were greater at lag 0–1, with ER values per IQR increase of 2.13, 1.73 and 1.79% (i.e., 0.44, 0.52 and 0.53% per 10ug/m³ increase). Consistently, Wu et al. reported that ER of mortality associated with IQR increase in PM₁₀ in Guangzhou was 1.77% (95%CI: 0.37–3.18%) in

2006–2009 and 2.03% (95%CI: 0.79–3.30%) in 2010–2013, and the effect estimates were very similar after adjusting for NO₂ or SO₂ [35]. The estimate of the PM₁₀ effect was also very similar to that reported in 90 American cities and in 38 Chinese cities (i.e., 0.5 and 0.44% per 10ug/m³ increase), while the effects of SO₂ and NO₂ we observed were smaller than the estimates in four Asian cities (i.e., 1.00 and 1.23% per 10ug/m³ increase) [34, 36].

The three-pollutant models produced abnormal health impact estimates of single pollutants (e.g., a negative effect of NO₂) and w19pt>The above phenomenon was due to strong collinearity among pollutants with

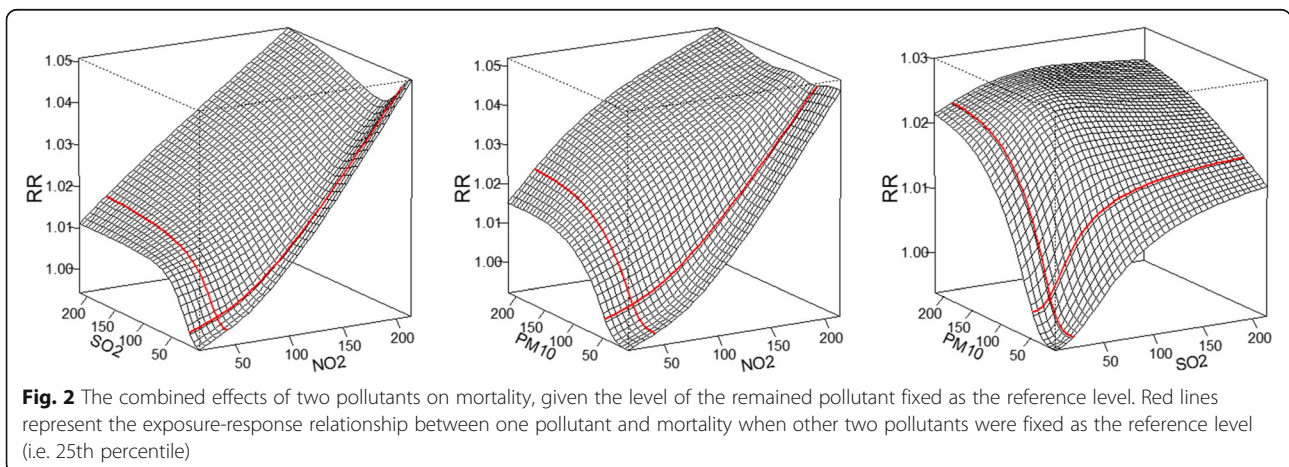


Table 2 Excess risk of mortality and 95% confidence intervals (%) associated with an IQR increment in air pollutant concentrations

Model ^a	PM ₁₀	NO ₂	SO ₂	Combined effects
I	2.03 (1.08–2.99)	1.63 (0.69–2.59)	1.86 (0.94–2.78)	5.63 (3.96–7.34)
II	1.98 (0.23–3.77)	− 1.24 (− 3.24–0.81)	1.47 (− 0.10–3.07)	2.20 (1.18–3.23)
III	1.35 (− 1.11–3.88)	− 0.49 (− 3.51–2.61)	2.46 (− 0.11–5.10)	2.78 (1.35–4.23)
IV	1.45 (− 0.33–3.27)	0.53 (− 1.42–2.52)	0.92 (− 0.65–2.52)	2.53 (1.03–4.01)

^aModel I, II, III and IV denote the single-pollutant model, the three-pollutant non-interaction model, the three-pollutant interaction model and the tensor product model, respectively

Spearman correlation coefficients of 0.60–0.83, which caused variance inflation of the regression parameters and bias in the statistical inference [11]. As to the combined effects of three pollutants, it was clear that the additive effect of a single pollutant based on the single-pollutant model was much higher than the combined effect estimated by other three multi-pollutant

models. The reasons are that the single pollutant model does not take into account the high correlation between pollutants, and one air pollutant may be a substitute for another air pollutant or the mixture of air pollutants [37].

The most innovative point of this study is that the tensor product quasi-Poisson regression model was, for the

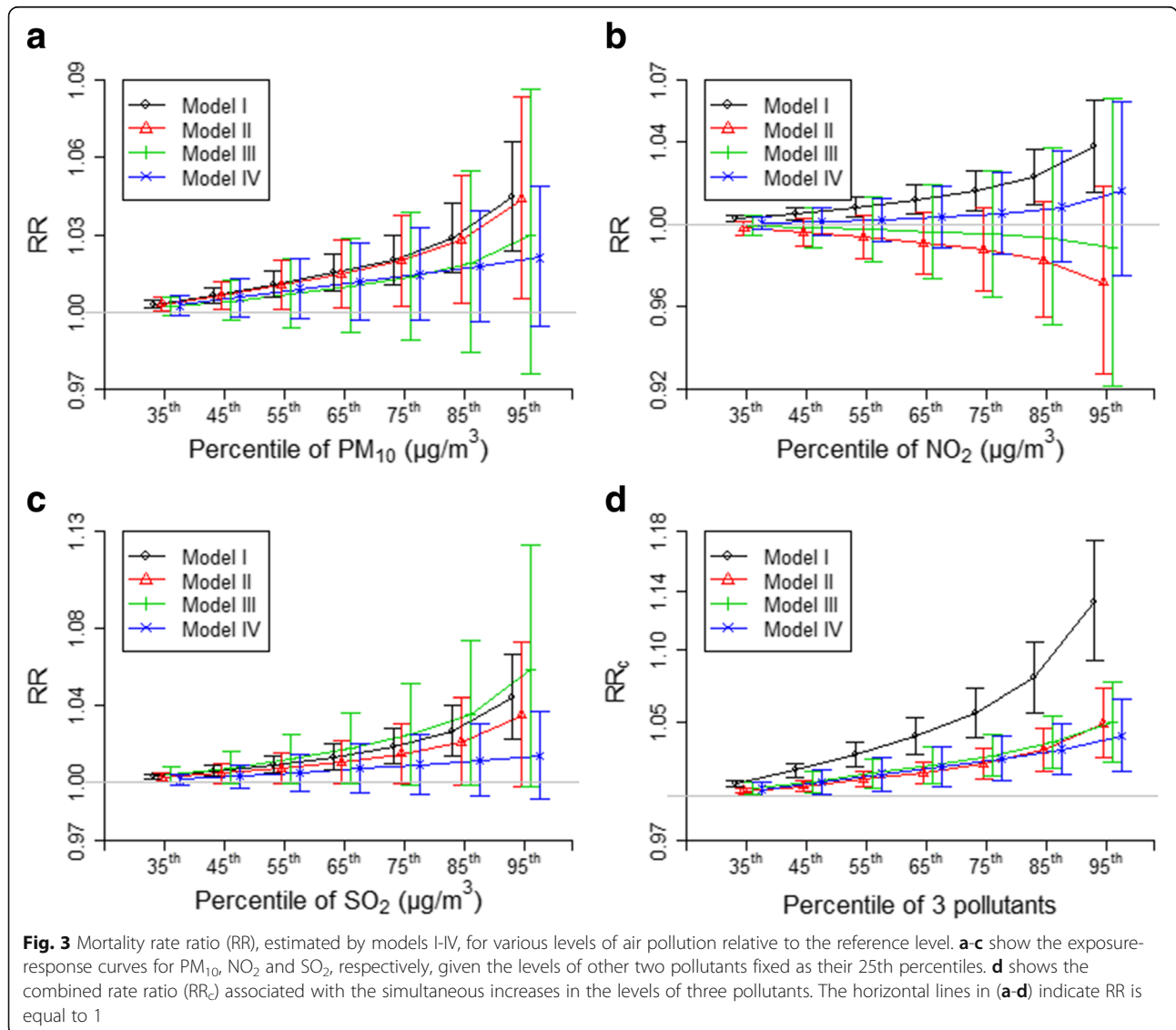


Table 3 The combined effects of three air pollutants on mortality by age, sex and educational attainment, estimated by the tensor product model

	Deaths (%)	ER ^a (95%CI)
Age (years)		
< 65 years	47,303(24.4)	-1.09 (- 2.81–0.66)
≥ 65 years	146,412(75.4)	3.71 (1.96–5.50)
Gender		
Female	84,519(43.6)	3.58 (1.47–5.74)
Male	109,916(56.4)	1.50 (0.05–2.99)
Educational attainment		
Low	170,629(92.6)	2.82 (1.19–4.49)
High	13,734(7.4)	2.19 (-1.11–5.60)

^aThe excess risk (ER) of all-cause mortality associated with an IQR increment of concentrations of three pollutants

first time, applied to assess the combined effect of several pollutants on adverse health outcome in the presence of multi-collinearity and to explore the complex interaction of several pollutants. Several methods that were proposed in the past made additional assumptions of the existence of a linear or specific functional relationship between atmospheric pollutants and the response variable or only explored the first order interaction; in contrast, the GAM with tensor product smooth not only relaxes those assumptions but also flexibly explores the nonlinear effect and the complex interaction of multiple pollutants. Specially, this method gives a highly visual depiction of the relationship between all-cause deaths and air pollutants, which may allow analysts to discover more undetectable characteristics of data than would be possible with the use of a simple model, and provide a new approach to comprehensive evaluation of health effects of multiple pollutants.

In this study, the tensor product model was also used to identify subpopulations that were more susceptible to exposure to a mixture of multiple pollutants. Some studies that assessed the health effect of single pollutants revealed that females were more susceptible than males to adverse effects of ambient air pollution [38]. Kan et al. observed that the effects of PM₁₀ on mortality were almost twice as strong in females [39]. The reasons for sexual differences were unclear and deserved to be further investigated. We observed greater combined effects of multiple pollutants in females than in males. Residents with low educational attainment were more sensitive to exposure to ambient air pollution than those with high educational attainment, which was consistent with previous studies [39, 40]. Most individuals with low educational attainment experience higher financial risk and, therefore, potentially also experience inferior living conditions and inadequate healthcare. Additionally,

individuals with low educational attainment were more likely to be occupationally exposed to air pollution [41].

Our study has some limitations. Although the GAM with tensor product smooth is flexible enough to explore the relationship between adverse health outcome and air pollutants, the “curse of dimensionality” occurs when the number of variables in the tensor product smooth increases, which is inherent in many flexible models [23]. The total dimensionality is the product of dimensionalities of each pollutant; thus, a large sample size is needed to accurately estimate parameters, and the computation is time-consuming. Moreover, it is difficult to interpret results of the model in high-dimensional spaces. We assessed health effects of three main pollutants to demonstrate the methods; however, other pollutants, such as ozone and PM_{2.5} that might be related to mortality were not considered in this study because of data unavailability during the study period.

Conclusions

Ambient air pollution has a significant impact on mortality. Our findings indicate that there exists a high correlation and a complex interaction among air pollutants. The single-pollutant model is unable to accurately estimate the combined effects of multiple pollutants. The simple multi-pollutant model with or without the first-order interaction term may yield misleading results when attributing health effects to individual pollutants. This study provides a novel method, called a tensor product quasi-Poisson model, for assessing the health effects of multiple pollutants, which can be applied in various epidemiologic studies to determine the combined effects of environmental, social and individual risk factors, especially if such factors are highly correlated and have nonlinear health effects.

Additional file

Additional file 1: FigureS1. Excess risk of mortality and 95% confidence intervals (%) associated with an IQR increment in moving average lag of 0–1 and 0–2 of air pollutants. Lag0-1and lag0-2 refer to moving average of the current day and preceding 1 day and 2 days. I-IV denote the models I-IV. (TIFF 551 kb)

Abbreviations

BKMR: Bayesian kernel machine regression; BMA: Bayesian model averaging; CI: Confidence interval; df: Degree of freedom; ER: Excess risk; GAM: Generalized additive model; HOD: Holiday; IQR: Interquartile-range; NO₂: Nitrogen dioxide; ns: Nature cubic spline; PACF: Partial autocorrelation function; PCA: Principal component analysis; PLSR: Partial least-square regression; PM₁₀: Particulate matter less than 10 μm in aerodynamic diameter; RR: Rate ratio; SO₂: Sulfur dioxide; te: Tensor product smooths

Acknowledgements

We appreciate Guangzhou Centre for Disease Control and Prevention and Guangzhou Bureau of Environmental Protection for providing all-cause death data and air pollution data, respectively.

Funding

This work was supported by National Nature Science Foundation of China [81573249] and Nature Science Foundation of Guangdong Province [2016A030313530]. The funders had no role in this study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

Raw data of all-cause mortality data, daily air pollution concentrations and daily metrological data were requested and obtained from Guangzhou Centre for Disease Control, Prevention and Guangzhou Bureau of Environmental Protection and China Meteorological Data Sharing Service System, respectively. Raw data will not be shared because the authors are not authorized for distribution of data.

Authors' contributions

CQ initiated the study. LL and SQ collected and cleaned data. LJ and SQ performed the statistical analysis. LJ and CQ drafted the manuscript. All authors made contributions to interpretation of data and revising the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 31 October 2018 Accepted: 29 March 2019

Published online: 24 April 2019

References

- Pope CA, Burnett RT, Krewski D, Jerrett M, Shi Y, Calle EE, et al. Cardiovascular mortality and exposure to airborne fine particulate matter and cigarette smoke shape of the exposure-response relationship. *Circulation*. 2009;120:941–8.
- Zanobetti A, Gold DR, Stone PH, Suh HH, Schwartz J, Coull BA, et al. Reduction in heart rate variability with traffic and air pollution in patients with coronary artery disease. *Environ Health Perspect*. 2010;118:324–30.
- Chen R, Kan H, Chen B, Huang W, Bai Z, Song G, et al. Association of particulate air pollution with daily mortality: the China air pollution and health effects study. *Am J Epidemiol*. 2012;175:1173–81.
- Hart JE, Bertrand KA, DuPre N, James P, Vieira VM, VoPham T, et al. Exposure to hazardous air pollutants and risk of incident breast cancer in the nurses' health study II. *Environ Health*. 2018;17:28.
- Dominici F, Burnett RT. Risk models for particulate air pollution. *J Toxicol Environ Health A*. 2003;66:1883–90.
- Liu Y, Xie S, Yu Q, Huo X, Ming X, Wang J, et al. Short-term effects of ambient air pollution on pediatric outpatient visits for respiratory diseases in Yichang city, China. *Environ Pollut*. 2017;227:116–24.
- Dominici F, McDermott A, Zeger SL, Samet JM. On the use of generalized additive models in time-series studies of air pollution and health. *Am J Epidemiol*. 2002;156:193–203.
- Vedal S, Brauer M, White R, Petkau J. Air pollution and daily mortality in a city with low levels of pollution. *Environ Health Perspect*. 2003;111:45–52.
- Jacoby DB. Synergism between rhinovirus infection and oxidant pollutant exposure enhances airway epithelial cell cytokine production. *Environ Health Perspect*. 2002;110:665–70.
- Vempilly J, Abejje B, Diep V, Gushiken M, Rawat M, Tyner TR. The synergetic effect of ambient PM_{2.5} exposure and rhinovirus infection in airway dysfunction in asthma: a pilot observational study from the Central Valley of California. *Exp Lung Res*. 2013;39:434–40.
- Francesca D, Peng RD, Barr CD, Bell ML. Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach. *Epidemiology*. 2010;21:187–94.
- Boezen HM, Sc VDZ, Postma DS, Vonk JM, Gerritsen J, Hoek G, et al. Effects of ambient air pollution on upper and lower respiratory symptoms and peak expiratory flow in children. *Lancet*. 1999;353:874–8.
- Winqvist A, Kirrane E, Klein M, et al. Joint effects of ambient air pollutants on pediatric asthma emergency department visits in Atlanta, 1998–2004. *Epidemiology*. 2014;25:666–73.
- Tango T. Effect of air pollution on lung cancer: a Poisson regression model based on vital statistics. *Environ Health Perspect*. 1994;102:41–5.
- Chen G, Song G, Jiang L, Zhang Y, Zhao N, Chen B, et al. Short-term effects of ambient gaseous pollutants and particulate matter on daily mortality in Shanghai, China. *J Occup Health*. 2008;50:41–7.
- Kan H, Chen B, Zhao N, London SJ, Song G, Chen G, et al. Part 1. A time-series study of ambient air pollution and daily mortality in Shanghai, China. *Res Rep Health Eff Inst*. 2010;154:17–78.
- Arif AA, Shah SM. Association between personal exposure to volatile organic compounds and asthma among US adult population. *Int Arch Occup Environ Health*. 2007;80:711–9.
- Holberg CJ, O'Rourke MK, Lebowitz MD. Multivariate analysis of ambient environmental factors and respiratory effects. *Int J Epidemiol*. 1987;16:399–410.
- Månsson K, Shukur G. A Poisson ridge regression estimator. *Econ Model*. 2011;28:1475–81.
- Koop G, Tole L. Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air? *J Environ Econ Manag*. 2004;47:30–54.
- Bobb JF, Valeri L, Claus HB, Christiani DC, Wright RO, Mazumdar M, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*. 2015;16:493–508.
- Sun Z, Tao Y, Shi L, Ferguson KK, Meeker JD, Kyun PS, et al. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. *Environ Health*. 2013;12:85.
- Wood SN. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*. 2006;62:1025–36.
- Zhao Y, Zhang Q, Levesley J. Multilevel sparse grids collocation for linear partial differential equations, with tensor product smooth basis functions. *Comput Math Appl*. 2018;75:883–99.
- Pulmannová S. Tensor products of quantum structures and their applications in quantum measurements. *Int J Theor Phys*. 2003;42:907–19.
- Li L, Yang J, Guo C, Chen PY, Ou CQ, Guo Y. Particulate matter modifies the magnitude and time course of the non-linear temperature-mortality association. *Environ Pollut*. 2015;196:423–30.
- Yang J, Ou CQ, Song YF, Li L, Chen PY, Liu QY. Estimating years of life lost from cardiovascular mortality related to air pollution in Guangzhou, China. *Sci Total Environ*. 2016;573:1566–72.
- Muggeo VM, Hajat S. Modelling the non-linear multiple-lag effects of ambient temperature on mortality in Santiago and Palermo: a constrained segmented distributed lag approach. *Occup Environ Med*. 2009;66:584–91.
- Braga A, Zanobetti A, Schwartz J. The time course of weather-related deaths. *Epidemiology*. 2001;12:662–7.
- Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*. 2006;295:1127–34.
- Guo Y, Li S, Zhang Y, Armstrong B, Jaakkola JJ, Tong S, et al. Extremely cold and hot temperatures increase the risk of ischaemic heart disease mortality: epidemiological evidence from China. *Heart*. 2013;99:195–203.
- Yang J, Liu HZ, Ou CQ, Lin GZ, Zhou Q, Shen GC, et al. Global climate change: impact of diurnal temperature range on mortality in Guangzhou, China. *Environ Pollut*. 2013;175:131–6.
- Guo Y, Tong S, Li S, Barnett AG, Yu W, Zhang Y, et al. Gaseous air pollution and emergency hospital visits for hypertension in Beijing, China: a time-stratified case-crossover study. *Environ Health*. 2010;9:57.
- Samet JM, Zeger SL, Dominici F, Currier F, Coursac I, Dockery DW, et al. The National Morbidity, mortality, and air pollution study. Part II: morbidity and mortality from air pollution in the United States. *Res Rep Health Eff Inst*. 2000;94:5–70 71–79.
- Rongshan W, Liuju Z, Xiaoliang H, Hongbing X, Shuo L, Baihuan F, et al. Temporal variations in ambient particulate matter reduction associated short-term mortality risks in Guangzhou, China: a time-series analysis (2006–2016). *Sci Total Environ*. 2018;645:491–8.

36. Yin P, He G, Fan M, Chiu KY, Fan M, Liu C, et al. Particulate air pollution and mortality in 38 of China's largest cities: time series analysis. *BMJ*. 2017;356:j667.
37. Roberts S. A new model for investigating the mortality effects of multiple air pollutants in air pollution mortality time-series studies. *J Toxicol Environ Health A*. 2006;69:417–35.
38. Clougherty JE. A growing role for gender analysis in air pollution epidemiology. *Cien Saude Colet*. 2011;16:2221–38.
39. Kan H, London SJ, Chen G, Zhang Y, Song G, Zhao N, et al. Season, sex, age, and education as modifiers of the effects of outdoor air pollution on daily mortality in Shanghai, China: the public health and air pollution in Asia (PAPA) study. *Environ Health Perspect*. 2008;116:1183–8.
40. Li L, Yang J, Song YF, Chen PY, Ou CQ. The burden of COPD mortality due to ambient air pollution in Guangzhou, China. *Sci Rep*. 2016;6:25900.
41. Finkelstein MM, Jerrett M, DeLuca P, Finkelstein N, Verma DK, Chapman K, et al. Relation between income, air pollution and mortality: a cohort study. *CMAJ*. 2003;169:397–402.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

