Atmospheric Environment 186 (2018) 136–143

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Association between multi-pollutant mixtures pollution and daily cardiovascular mortality: An exploration of exposure-response relationship

Yuanren Tong[a,1], Kai Luo[b,c,1], Runkui Li[d,e], Lu Pei[b,c], Ang Li[b,c], Mingan Yang[f], Qun Xu[b,c,*]

[a] Chinese Academy of Medical Sciences & Peking Union Medical College Beijing 100005, China
[b] Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China
[c] Center of Environmental and Health Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100005, China
[d] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
[e] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Science, Beijing 100101, China
[f] Division of Biostat & Epidemiology, Graduate School of Public Health, San Diego State University, San Diego, Ca 92182, USA

ARTICLE INFO

Keywords:
Air pollution
Bayesian kernel machine regression
Exposure-relationship
Cardiovascular mortality

ABSTRACT

Evidence of combined mortality effects of multi-pollutant on cardiovascular diseases (CVD) and the corresponding exposure-response (ER) relationship is limited. In this paper, we examined the association between four ambient air pollutants (i.e., fine particulate matter, $PM_{2.5}$, particulate matter, PM_{10}, nitrogen dioxide, NO_2, and sulfur dioxide, SO_2.) and CVD mortality and the corresponding ER relationship after incorporating the potential interaction among the multiple pollutants. Bayesian kernel machine regression (BKMR) was used to evaluate the ER relationship and to explore the interactions between pollutants. The results showed that PM_{10} and SO_2 were dominant pollutants from 0 to 2 days, while $PM_{2.5}$ and NO_2 had strong effect on CVD mortality from 3 to 4 days. Generally, $PM_{2.5}$ and NO_2 had the similar ER relationship across different moving average concerning the CVD mortality. For the interaction among the multiple pollutants, we found that there is no interaction between particle pollutants (i.e. $PM_{2.5}$ and PM_{10}) and gaseous pollutants (i.e.NO_2 and SO_2). On the contrary, there might be an interaction between $PM_{2.5}$ and PM_{10} though this interaction was detected by visually comparing the slopes of ER curves of a given particle pollutant at different levels of the other particle pollutant. But there is a lack of statistical significance test for this interaction. This study suggests that different ambient air pollutants might have the dominant effect on CVD deaths during different moving average, though there might not be statistical significant interactions among the ambient air pollutants in present study.

1. Introduction

It is well known that ambient air pollution is associated with a variety of adverse health outcomes, ranging from preclinical changes to death (Di et al., 2017; Mustafic et al., 2012; Rich et al., 2012; Sarnat et al., 2001). The latest estimated deaths attributed to ambient fine particulate matter were approximate 4.2 million worldwide and 1.1 million in China in 2015 (Cohen et al., 2017). A recent simulation study estimated that 241,000 life-years would benefit from the reduction in concentration of ambient $PM_{2.5}$ that reached the level during 2008 Beijing Olympic period (Huang et al., 2017). These studies have greatly promoted the understanding of the effects of individual pollutant but were mainly based on the epidemiologic studies of single pollutant, where effect estimations were obtained from models with single pollutant or adjusted for col-pollutant in the same statistical model.

Although it is essential to understand the effects of exposure to a single pollutant, people are invariably exposed to a multi-pollutant mixture instead of a single pollutant (Braun et al., 2016). Diseases related to ambient air pollution are rarely attributed to a single pollutant, the contribution of multi-pollutant would be more important and significant. Thus, it is important to study the effects of multi-pollutant exposure to elaborate its hazard health effects and to provide scientific evidence for air pollution control policy (Dominici et al., 2010). Actually, the unknown interaction among air pollutants have made it a daunting task to estimate the effects of the multi-pollutant exposure on health (Davalos et al., 2017). Some statistical methods exploring multi-

* Corresponding author. Department of Epidemiology and Biostatistics, Institute of Basic Medicine Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China.
E-mail address: xuqun@ibms.cams.cn (Q. Xu).
[1] Authors contribute equally to this work.

https://doi.org/10.1016/j.atmosenv.2018.05.034
Received 23 October 2017; Received in revised form 14 May 2018; Accepted 15 May 2018
Available online 24 May 2018

pollutant adverse health effects have been proposed. For instance, Schildcrount et al. (Schildcrout et al., 2006) proposed an additive effect model with decomposed effects to examine the combined effect of a simultaneous increase in air pollution concentrations on asthma exacerbations. But this approach is highly dependent on the selected models. Other researchers have proposed approaches based on dimension reduction techniques such as factor analysis, principle component analysis, latent class analysis, to allow for the correlation and additive/motive interaction of air pollutants (Pachon et al., 2012; Roberts, 2006; Sacks et al., 2012; Yang et al., 2013; Zanobetti et al., 2014). However, all these models are not flexible in assessing the interaction of different air pollutants and the corresponding combined effects, particularly for the potential nonlinear exposure response relationship.

Recently, Bobb et al. (2015) proposed a statistical learning method, based on Bayesian kernel machine regression (BKMR), to estimate joint effects of multiple pollutants, while allowing for potential nonlinear or nonadditive associations between a given pollutant and health outcome of interest. This approach has some appealing advantages in estimating health effects caused by multi-pollutant and identifying the dominant pollutants that have the strongest effect on the health outcome under certain condition (Bobb et al., 2015; Coull et al., 2015). Moreover, the complicated nonlinear exposure-response(ER) relationship can be easily estimated with this method.

Given the limited evidence of effects of multi-pollutant exposure on daily mortality and the corresponding ER relationships, in present study, we adopted this approach to identify the ambient air pollutants that were most closely associated with daily cardiovascular mortality, and to calculate the ER relationship between air pollutants and cardiovascular mortality while incorporating the interaction of multiple pollutants.

2. Materials and method

2.1. Analysis plan

This study was comprised of two parts: simulation study for the selection of prior distributions of parameters of BKMR and its application. We presented the detailed introduction of simulation in supplementary materials online for conciseness. The implementation of BKMR was based on the information generated from the simulation study. The flow path of BKMR modeling in this study was demonstrated in S.Fig. 1 in supplementary materials.

2.2. Data collection

Data of daily mortality caused by cardiovascular disease (ICD-10, I00-I99) from 2008 to 2011 was retrieved from Causes of Death Registry of Chinese Center for Disease Control and Prevention (China CDC). Data of daily air pollutants, including PM_{10}, NO_2 and SO_2, was collected from 12 national air quality monitoring stations (NAQM) in Beijing during the same period. To evaluate the association between $PM_{2.5}$ and CVD mortality and the ER relationship, we also obtained the daily $PM_{2.5}$. Because $PM_{2.5}$ was not yet routinely monitored in China until late December 2012, data of daily $PM_{2.5}$ during the same period were obtained from the air quality monitoring station of the U.S. Embassy in China, which is located in the Chaoyang district of Beijing. Meteorological data, including daily mean temperature and relative humidity were retrieved from the Beijing Observatory (Station No.: 54511) of the China Meteorological Administration, which is located in the Daxing district in Beijing. Table 1 shows the descriptive statistics of daily death counts and ambient air pollutants in Beijing, China.

In order to explicitly interpret results of multiple pollutants within the BKMR framework, data of ambient pollutants were first standardized before implementing BKMR.

Table 1
Descriptive statistics of daily death counts and air pollutants used in the application.

	Mean	SD	Min	P25	P50	P75	Max
Pollutants							
NO_2(μg/m^3)	49.82	23.13	5.37	33.17	45.90	60.00	180.67
PM_{10}(μg/m^3)	117.75	74.14	4.91	65.09	104.00	146.55	651.18
$PM_{2.5}$(μg/m^3)	95.68	70.83	5.83	41.79	80.09	127.92	492.75
SO_2(μg/m^3)	32.27	31.66	3.00	10.96	20.00	42.00	202.00
Death Count							
CVD	99.57	20.36	54.00	85.00	97.00	113.00	173.00

*P25, P50, P75 refer to 25th, 50^{th} and 75th percentile, respectively.

2.3. Statistical analysis

To accommodate for the potential synergistic and nonlinear effects among multiple air pollutants, we used BKMR, a novel approach for multi-pollutant mixtures that flexibly models the joint effect of mixtures using a kernel function (Bobb et al., 2015). This method allows estimation of nonlinear or nonadditive ER function for a set of correlated pollutants accounting for uncertainty.

Specifically, kernel machine regression is a data processing which is also known as kernel trick. The process is a mapping from the low-dimension data space to a new high-dimension space. The link of the two spaces is the kernel function. In this study we used the Gaussian kernel function because it is available for general conditions. However, it is valuable to point out that according to Mercer's theorem, each positive semi-definite function can perform as a kernel function. The general Gaussian kernel function is shown below as formula (1):

$$K(\mathbf{z},\mathbf{z}') = e^{\left\{-\frac{1}{\rho}\sum_{m=1}^{M}(z_m - z_m')^2\right\}} \tag{1}$$

where $\mathbf{z} = (z_1,...,z_M)^T$ and $\mathbf{z}' = (z_1',...,z_M')^T$ are the two $M \times 1$ vectors representing two pollutants' concentration of M days, respectively.

Posterior inclusion probabilities (PIPs) are calculated to conduct the variable selection process (Coull et al., 2015), with larger PIPs of a pollutant indicating the stronger association with health outcome.

To specify the model, we use $\mathbf{pol} = (\mathbf{c(CO)},\ \mathbf{c(NO_2)},\ \mathbf{c(NO_x)},\ \mathbf{c(PM_{10})})$ as the matrix of the variables. For the convenience of expression here, we present $(\mathbf{c(CO)},\ \mathbf{c(NO_2)},\ \mathbf{c(NO_x)},\ \mathbf{c(PM_{10})})$ as $(\mathbf{z_1,z_2,z_3,z_4})$, while $\mathbf{z_i}$ was a 1461×1 vectors containing the corresponding pollutant's concentration through 1461 days (four years, 2008–2011) in the dataset. We defined $h(\mathbf{pol}) = \sum_{i,j=1}^{4} K(\boldsymbol{z_i},\boldsymbol{z_j})$ while $K(\boldsymbol{z_i},\boldsymbol{z_j})$ is the Gaussian kernel shown in formula (1). The model can be written as:

$$\log(E(deaths)) = h(c(CO), c(NO_2), c(NO_x), C(PM_{10})) + \boldsymbol{X_i^T\beta} \tag{2}$$

where X_i is a vector contains potential confounders and here X_i is the meteorological conditions. The prior distribution of β is set as the defaulted flat distribution.

The selection of the prior distribution is important for BKMR. It is advised to avoid overfitting or underfitting, which can easily cause oscillations, in the environmental health studies. Following this line, we performed a simulation study, which was demonstrated in supplementary materials.

We first conducted a simulation study in this experiment to study the performance of BKMR in the air pollution exposure-response researches. The simulation study also helped us to find a proper prior distribution in BKMR. Details of the simulation study can be found in the supplementary materials.

Given the potential lagged effect of ambient air pollutants on daily CVD mortality, we fit BKMR with 0–1, 0–2, 0–3, 0–4 days (MV01-04) moving average, for the cumulative effects of air pollutants. The curves of ER relationship and PIPs were fitted and calculated in the process of