Quantification of the Exposure–Lag–Response Association Between Air Pollution and Respiratory Disease Morbidity in Chongqing City, China

Chenxi Xia¹ · Jingdong Ma¹ · Jing Wang¹ · Jing Huang² · Qing Shen² · Yuanlan Chen² · Youfan Jiang³

Received: 20 May 2017 / Accepted: 28 June 2018 / Published online: 6 July 2018
© The Author(s) 2018

Abstract

The health effects of air pollution on respiratory disease morbidity and mortality commonly vary with air pollutants' intensity and latency after exposure. However, the exposure–lag–response association is seldom explored, especially in China, where the smog has led to public displays of dissatisfaction and became a big political issue. In this study, the distributed lag nonlinear model (DLNM) and the generalized additive model (GAM) was used to estimate the relative risk of air pollution exposure history for respiratory disease admission in Chongqing City, China. The results reveals that the health effect of inhalable particulate matter (PM₁₀) can be estimated using a cross-basis by combining two linear functions that respectively model the exposure–response curve and lag–response curve. The risks of sulfur dioxide (SO₂) and nitrogen dioxide (NO₂) exposure are insignificant. The relative risk of PM₁₀ exposure decreases as time goes by. The greater PM₁₀ intensity, the more quickly the relative risk decreases. The relative risk is higher than 1.0 in 0 to 23 days after exposure. The greater PM₁₀ intensity, the higher the relative risk is. On the contrary, the relative risk is lower than 1.0 after 23 days. The greater PM₁₀ intensity, the lower the relative risk is. This study clarified the health effects of air pollution on respiratory disease admissions and can be used to predict the relative risk of air pollution or the cumulative health effect of a history of air pollution exposure in Chongqing City.

Keywords Air pollution · Respiratory disease · Distributed lag nonlinear model · Generalized additive model · Lag range

1 Introduction

Several studies have demonstrated the adverse effects of air pollution on morbidity and mortality [1–3]. The daily counts of patients can be described as the result of a series of air pollution exposure events of different intensities for a few days. The health outcomes of air pollution exposure at a given day may be sustained in the future and may vary according to the intensity and the lag period. Gasparrini [4] referred to this type of dependency as the exposure–lag–response association.

The distributed lag nonlinear model (DLNM) [4–6] was developed to quantify the effect. The model is based on the

definition of a cross-basis, which is obtained by combining of two linear or nonlinear functions to model the exposure–response and lag–response relationships, respectively. Researchers commonly use DLNM combining with the generalized linear model (GLM) [7] or the generalized additive model (GAM) [8]. With this modeling framework, the main complexity of quantifying the association lies in model selection, including the selection among various candidates of confounders, smoothing functions, and parameters of the GAM smooth terms, as well as the lag–response functions, exposure–response functions, and parameters of the DLNM cross-basis. Sensitivity analysis is also required when a priori options are used in the models.

To our knowledge, the effect of air pollution on respiratory disease morbidity is seldom quantified with DLNM, especially in China, which is suffering from smog pollution severely [9]. Smog pollution has led to public displays of dissatisfaction and became a big political issue. According to the 2014 Chinese Environmental State Report published by the Ministry of Environmental Protection of China, only eight of the 74 monitored cities met urban air quality standards in 2014 in terms of annual average value, accounting for 10.8%. In other words, 89.2% of the monitored cities experienced air pollution at a range of different levels.

✉ Jing Huang
huangjing527@hotmail.com

¹ School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China

² Department of Respiratory Medicine, Chongqing General Hospital, Chongqing, China

³ Department of Respiratory Medicine, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China

In this paper, we aim to quantify the association between air pollution and daily respiratory disease admissions in Chongqing City. The most significant confounders are identified by model selection. The relative risk of air pollution is estimated with DLNM. The result sensitivity to DLNM functions and parameters, especially the lag range, is investigated.

2 Data

Chongqing City is the capital of Chongqing province in China, with an area of 5473 km² and a population of 8.18 million in 2014. Located in southwest China, the city is between the Qinghai Tibet Plateau and the Middle-Lower Yangtze plains, with complex geographic features of mountains and hills. Chongqing is also called the city of hills. As a well-known “stove,” summers in Chongqing are hot, with temperatures between 27 and 38 °C in July and August, and a recorded maximum temperature of 43.8 °C. Chongqing is a foggy city, with an average of 104 foggy days per year, more than London and Tokyo.

A total of 8841 patients with respiratory diseases were hospitalized at two top level hospitals in Chongqing City from January 1, 2008 to December 31, 2013. One hospital is the Chongqing General Hospital, which has 736 beds and admitted 19,713 patients in 2013. The other hospital is the Second Affiliated Hospital of Chongqing Medical University, which has 1380 beds and provided service to approximately 45,000 inpatients in 2013. Patients are selected according to their primary diagnosis ICD-10 codes in electronic medical record (EMR). All respiratory diseases appearing as inpatient primary diagnosis except tumor are used, including pneumonia, rhinitis, sinusitis, bronchitis, chronic obstructive pulmonary disease, and asthma. ICD-10 codes used are A15, A36, A37, J00-J06, J11, J12, J15, J18, J20, J21, J30-J32, J40-J46, J67, and J82.

Daily meteorological data from January 1, 2008 to December 31, 2013 is obtained from the Chongqing Meteorological Bureau. The data includes the minimum, maximum, and mean temperature; the maximum wind speed; the minimum, maximum, and mean atmospheric pressure; and the mean relative humidity.

Daily air pollution data for the same period was collected by the Chongqing Environmental Protection Bureau. Daily air pollution intensities, including the intensities of inhalable particulate matter (PM₁₀), sulfur dioxide (SO₂), and nitrogen dioxide (NO₂), are collected at 29 monitoring stations distributed across the city and averaged. The actual levels of air pollution intensity are not published. Instead, the intensity of each air pollutant is represented as an air pollution indicator (API), which is calculated with piecewise interpolation as following equation:

$$API = \frac{I - I_i}{I_{i+1} - I_i} (API_{i+1} - API_i) + API_i. \quad (1)$$

In Eq. (1), I represents air pollutant intensity. I_i and API_i are parameters shown in Table 1, which are defined in the ambient air quality standards of China (GB 3095-1996) [10].

The daily meteorological and air pollution data have no missing values. A summary of the data is provided in Table 2.

3 Methods

A Poisson distribution and a log link are used during fitting. This study uses GAM and DLNM to estimate the effect of air pollution on daily admissions in the following two steps.

At first, confounders are controlled in a forward stepwise search style. All candidate confounders, as listed in Table 3, are added into the model iteratively. In each iteration, the best confounder, which minimizes the Akaike's information criterion (AIC) and Bayesian information criterion (BIC) of the model, is selected. A confounder or its moving averages will not be used since it or one of its moving averages has been selected. In other words, a confounder and its moving average terms are never simultaneously added to a model. The value of n that fitted the data best for each meteorological data item is set by model selection. When all candidate confounders are used or no confounder can decrease the AIC and BIC of the model, we get a baseline model, shown as Eq. (2):

$$\begin{aligned} \text{Log}(E(y_t)) = & \alpha + \beta_1 \text{Hol}_t + \beta_2 \text{DOW}_t + s(t) \\ & + s(\text{Max_Temp}_{t,8}) + s(\text{Hum}_t). \end{aligned} \quad (2)$$

In Eq. (2), t is the date sequence. y_t refers to the number of the observation. $E(y_t)$ denotes the estimated daily hospital admissions on day t . α is the intercept. Hol_t denotes the holiday indicator. DOW_t is the day of week indicator. The last three terms are the smooth terms of the selected confounders. All smooth terms are defined by a thin plate regression spline. The smoothing parameters are automatically selected with generalized cross validation (GCV). The term $s(t)$ is the smooth term of date sequence. The term $s(\text{Max_Temp}_{t,8})$ is the smooth term of the last 8-day moving average of maximum temperature. The term $s(\text{Hum}_t)$ is the smooth term of mean relative humidity. The descriptive statistics of the original data of maximum temperature and mean relative humidity is shown in Table 2. Equation (2) is a generalized additive model

Table 1 API standard

Air pollutant	$i = 1$		$i = 2$		$i = 3$	
	API ₁	I_1 (mg/m ³)	API ₂	I_2 (mg/m ³)	API ₃	I_3 (mg/m ³)
PM ₁₀	50	0.05	100	0.15	200	0.25
NO ₂	50	0.08	100	—	200	0.12
SO ₂	50	0.05	100	0.15	200	0.25

Table 2 Descriptive statistics of the daily respiratory disease admission, meteorological, and air pollution data of Chongqing City from Jan 1, 2008 to Dec 31, 2013

	Mean $\pm$ SD	Minimum	25th quartile	Median	75th quartile	Maximum
Daily admissions	4.04 $\pm$ 3.35	0	2	3	6	23
Minimum temperature (°C)	16.16 $\pm$ 7.19	0.40	10.00	16.80	22.00	32.00
Maximum temperature (°C)	22.74 $\pm$ 9.30	2.40	15.20	22.70	30.20	42.70
Mean temperature (°C)	18.94 $\pm$ 8.04	1.70	12.00	19.10	25.20	36.90
Maximum wind speed (m/s)	3.01 $\pm$ 1.02	1.20	2.30	2.80	3.40	10.90
Minimum atmospheric pressure (hPa)	979.92 $\pm$ 8.53	955.10	972.60	980.00	986.47	1003.20
Maximum atmospheric pressure (hPa)	985.94 $\pm$ 9.14	964.70	978.00	986.20	992.90	1061.80
Mean atmospheric pressure (hPa)	983.05 $\pm$ 8.62	961.50	975.60	983.20	989.70	1005.70
Mean relative humidity (%)	74.90 $\pm$ 13.36	25	66	76	85	100
PM ₁₀ (API)	74.42 $\pm$ 27.26	10.00	56.00	71.00	88.75	194.00
NO ₂ (API)	27.48 $\pm$ 13.15	0	19	24	33	89
SO ₂ (API)	41.15 $\pm$ 17.47	0	29	39	53	103

(GAM) with five explanatory variables, which is often combined with a distributed lag nonlinear model (DLNM) to explore the nonlinear exposure–lag–response associations, as in [4, 5, 12].

Then, a cross-basis matrix obtained by applying DLNM to air pollutant intensity is added to the right side of the baseline model. For air pollutant a_k ($a_k \in \{\text{PM}_{10}, \text{NO}_2, \text{SO}_2\}$), the cross-basis matrix $M(f(a_{k,t}), w(l))$ is a combination of the

exposure–response function $f(a_{k,t})$ and the lag–response function $w(l)$, where $a_{k,t}$ is the observed intensity of a_k at day t , and $0 < l \leq L$. L is the lag range, which is set a priori to 40. The candidates of $f(a_{k,t})$ are linear and B-spline functions. The knots for the B-spline of $f(a_{k,t})$ are placed at 25th, 50th, and 75th quartiles of $a_{k,t}$. The candidates of $w(l)$ are piecewise constant, linear functions, and B-spline functions with or without intercepts. The knots for the B-spline of $w(l)$ are placed at 1/3 and 2/3 of L . The B-spline without an intercept will left-constrain the smooth lag–response curve to start from 0. This assumption implies that the air pollution exposure at a given day does not exhibit health effects for that day. A total of 10 models are tested for each a_k . And the best model with minimum AIC and BIC is selected to estimate the health effects of the air pollutant.

To test the sensitivity of DLNM to different parameters, we change the lag range L , B-spline degrees, and knots when building the cross-basis matrix $M(f(a_{k,t}), w(l))$. New models with same parameter configurations are tested by their AIC and BIC. Best-fitted models for each configuration are selected and compared with each other.

All modeling and analysis are implemented with R (version 3.0.2), especially with R packages of mgcv (version 1.8-6) and dlnm (version 2.1.3).

4 Results

The best-fitted baseline model is represented by Eq. (2). The confounders and their coefficients and significance of terms are listed in Table 4. The AIC and BIC of the baseline model are 9153.037 and 9568.598, respectively. The significant confounders are date sequence day of the week indicator, holiday indicator, maximum temperature of the previous 8 days, and mean relative humidity.

Table 3 All candidate confounders

Code	Description
t	Date sequence
DOW_t	Day of the week indicator
Hol_t	Holiday indicator
Season_t	Seasonal indicator
$\text{Min_Temp}_{t,n}$	Minimum temperature and its moving averages for the last n days
$\text{Max_Temp}_{t,n}$	Maximum temperature and its moving averages for the last n days
$\text{Mean_Temp}_{t,n}$	Mean temperature and its moving averages for the last n days
$\text{Hum}_{t,n}$	Mean relative humidity and its moving averages for the last n days
$\text{Win}_{t,n}$	Maximum wind speed and its moving averages for the last n days
$\text{Min_P}_{t,n}$	Minimum pressure and its moving averages for the last n days
$\text{Max_P}_{t,n}$	Maximum pressure and its moving averages for the last n days
$\text{Mean_P}_{t,n}$	Mean atmospheric pressure and its moving averages for the last n days

$t = 1, 2, \dots, 2186$. $n = 1, 2, \dots, 15$. Each year is split into two seasons of summer and winter. The summer includes the months from May to October because it contains most of the days with a maximum temperature above 30 °C in the study period (543 of total 569 days, 95.43%)

Table 4 Coefficients and significance of baseline model terms

Coefficients and significance of linear terms					
	Estimate	Std. error	Z value	Pr (> z)	Significance code
α	1.412789	0.054242	26.046	< 2e-16	***
Hol _t :Y	-0.578969	0.047480	-12.194	< 2e-16	***
DOW _t :2	0.222753	0.057262	3.890	0.0001	***
DOW _t :3	0.002573	0.058828	0.044	0.9651	
DOW _t :4	-0.025061	0.059054	-0.424	0.6713	
DOW _t :5	-0.061065	0.059363	-1.029	0.3036	
DOW _t :6	-0.151399	0.059809	-2.531	0.0114	*
DOW _t :7	-0.058303	0.048819	-1.194	0.2324	
Coefficients and approximate significance of smooth terms					
	Estimated degrees of freedom	χ^2	p value	Significance code	
$S(t)$	55.935	1884.80	< 2e-16	***	
$S(\text{Max_Temp}_{t,8})$	2.538	17.96	0.000634	***	
$S(\text{Hum}_t)$	6.604	22.58	0.001939	**	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

The cross-basis matrixes of DLNM are added into the baseline model to quantify the health effects of air pollution exposure. Table 5 presents models with cross-basis from the combination of different $f(a_{k,t})$ and $w(l)$. The fit results for PM₁₀ are expressed by AIC and BIC. Model 2 achieves the best performance.

Relative risk (RR) of respiratory disease admissions associated with PM₁₀ exposure is calculated versus PM₁₀ API=0. As shown in Fig. 1, this model indicates a linear decrease of the relative risk along lag days for all PM₁₀ intensities. The greater PM₁₀ intensity, the more quickly the relative risk decreases. The relative risk is higher than 1.0 in 0 to 23 days after exposure. The greater PM₁₀ intensity, the higher the relative risk is. On the contrary, the relative risk is lower than 1.0 after 23 days. The greater PM₁₀ intensity, the lower the relative risk is.

Some results of the sensitivity analysis are shown in Table 6. The best-fitted models are consistent across different functions and parameters used to fit the cross-basis. This trend suggests that DLNM is robust in this aspect for our study design. When we change the lag range to 45 days or greater, the fit results become unstable. This trend suggests that DLNM is sensitive to a lag range greater than 40 days for our study design.

The fit results of SO₂ and NO₂ are unstable for a lag range of 5–40 days. The t tests show that the linear or smooth terms of SO₂ and NO₂ on each exposure day are insignificant (results not shown).

A Q-Q plot of the deviance residuals of the model versus a standard normal distribution is shown in Fig. 2, which shows that the model fits the data very well. The autocorrelation plots of the raw data and the deviance residuals are shown in Figs. 3

Table 5 Models with different combination of $f(a_{k,t})$ and $w(l)$, with the AIC and BIC values of alternative models to estimate the exposure-lag-response association between PM₁₀ and daily admissions

Model	$f(a_{k,t})$	$w(l)$	Degrees of freedom	AIC	BIC
1	Linear	Constant	1	9043.722	9450.228
2	Linear	Linear	2	<i>9038.160</i>	<i>9441.918</i>
3	Linear	Piecewise constant #	3	9039.637	9448.764
4	Linear	B-spline **	5	9041.625	9462.108
5	Linear	B-spline, no intercept **	4	9041.641	9458.438
6	B-spline, no intercept *	Constant	5	9047.686	9462.447
7	B-spline, no intercept *	Linear	10	9046.895	9491.625
8	B-spline, no intercept *	Piecewise constant #	15	9050.277	9513.919
9	B-spline, no intercept *	B-spline **	25	9058.382	9585.262
10	B-spline, no intercept *	B-spline, no intercept **	20	9053.113	9554.522

The italicized values are the smallest AIC and BIC. It shows that the corresponding model fits the data best
Lag range L of $w(l)$ is 40

Cut-offs at 13.33 and 26.67 lag

*Knot at 56.00, 71.00, and 88.75 API, degree is 2

**Knot at 13.33 and 26.67 lag days, degree is 2

Fig. 1 Relative risk (RR) of respiratory disease admissions associated with PM₁₀ exposure of 0–200 API and a lag period of 0–40 days. (1.1) 3-D surface of the exposure–lag–response association estimated from model 2. (1.2) Lag–response curves for PM₁₀ exposure of 30, 80, 100, 130, and 180 API. (1.3) Exposure–response curves at lag of 5, 15, 20, 25,

and 35 days. The two boldface lines in (1.1) correspond to the dotted lines (1.2) and (1.3), which represent the lag–response curve for a given PM₁₀ level of 100 API and the exposure–response curve for lag 20, respectively. Relative risk is calculated versus PM₁₀ API = 0

and 4 separately. Although the raw data plot shows strong positive autocorrelation, the residuals plot shows that the model results in random residuals. The result of Ljung-Box test also supports the adequacy of the model. For significance level 0.05 and lags number 40, the test statistic is 55.704, less than $\chi^2_{1-0.05,40}$, which is 55.758, and p value is 0.051.

5 Discussion

In this paper, we use DLNM and GAM to estimate the risk of air pollution exposure for respiratory disease admission in Chongqing City, China. The results suggest that the exposure–lag–response association between the PM₁₀ intensity and

Table 6 AIC and BIC values of models with different parameter configurations

$L = 40$	$D = 3, k1 = 3, k2 = 2$		$D = 4, k1 = 3, k2 = 2$		$D = 2, k1 = 2, k2 = 1$		$D = 2, k1 = 4, k2 = 3$	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Model 1	9043.722	9450.228	9043.722	9450.228	9043.722	9450.228	9043.722	9450.228
Model 2	<i>9038.160</i>	<i>9441.918</i>	<i>9038.160</i>	<i>9441.918</i>	<i>9038.160</i>	<i>9441.918</i>	<i>9038.160</i>	<i>9441.918</i>
Model 3	9039.637	9448.764	9039.637	9448.764	9038.326	9442.467	9041.529	9456.288
Model 4	9043.093	9469.047	9044.841	9476.372	9041.262	9456.026	9043.290	9469.314
Model 5	9042.948	9463.984	9044.491	9470.767	9040.549	9451.628	9043.266	9464.784
Model 6	9043.679	9462.258	9043.425	9474.670	9046.302	9464.442	9046.116	9462.969
Model 7	9042.685	9486.532	9042.628	9511.374	9042.673	9470.386	9042.485	9487.156
Model 8	9045.366	9522.992	9049.272	9556.254	9042.926	9470.199	9050.529	9562.277
Model 9	9060.801	9637.438	9065.786	9727.880	9051.187	9513.823	9060.457	9639.693
Model 10	9054.640	9573.041	9054.419	9682.717	9047.263	9495.915	9052.308	9597.479
$D = 2, k1 = 3, k2 = 2$	$L = 35$		$L = 45$		$L = 50$		$L = 55$	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Model 1	9058.979	9466.291	9025.241	9431.648	9007.285	9399.318	8994.243	9377.228
Model 2	<i>9054.500</i>	<i>9435.352</i>	<i>9023.086</i>	9429.788	9006.947	9400.759	8992.510	9379.772
Model 3	9056.833	9443.104	9024.677	9435.308	9007.931	<i>9398.131</i>	<i>8991.109</i>	<i>9375.012</i>
Model 4	9059.656	9457.632	9027.151	<i>9423.627</i>	9009.690	9407.571	8994.050	9390.293
Model 5	9059.606	9452.563	9024.875	9440.790	9009.830	9401.585	8993.400	9385.060
Model 6	9060.924	9480.547	9027.231	9435.852	9003.527	9443.014	8996.513	9422.528
Model 7	9058.013	9505.738	9029.384	9465.807	<i>9003.499</i>	9471.386	8996.428	9453.938
Model 8	9059.733	9533.770	9031.459	9486.872	9010.385	9481.789	8994.547	9463.954
Model 9	9075.268	9604.143	9044.552	9559.113	9019.209	9506.114	9003.211	9491.895
Model 10	9068.687	9568.224	9038.638	9526.136	9013.408	9474.671	8994.807	9481.404

The italicized values are the smallest AIC and BIC. It shows that the corresponding model fits the data best

L is the lag range. D is the B-spline degree of both $f(a_k, \cdot)$ and $w(l)$. $k1$ is the number of B-spline knots for $f(a_k, \cdot)$. $k2$ is the number of B-spline knots and piecewise constant cut-offs for $w(l)$. Knots and cut-offs are evenly distributed in the range of PM_{10} intensity and lag range. We varied L to 35–60, D to 2–4, $k1$ to 2–4, and $k2$ to 1–3 (some results not shown)

Fig. 2 Q-Q plot of the residuals of the best-fitted model

the risk can be represented by a bi-dimensional cross-basis. This cross-basis combines two linear functions. One models the exposure–response curve along the PM_{10} intensity and the other models the lag–response curve along the lag days. The results do not suggest a direct linear relationship between the relative risk and the PM_{10} intensity or PM_{10} exposure lag days because the coefficients of the two linear functions are different at different PM_{10} intensities or lag days. The respiratory disease morbidity surface is skewed as shown in Fig. 1(1.1).

A basic assumption of DLNM is that the exposure–response and lag–response functions are independent [4]. This assumption is not supported by our results because the coefficients of the exposure–response function change at different lag days and the coefficients of the lag–response function change with different PM_{10} intensities. This trend may cause the undercoverage of confidence intervals and inflated type I errors for tests, as suggested by Gasparini [4].

Sensitivity analysis shows that the fit with the best performance is robust when different functions and parameters are used to model the exposure–lag–response surface. However, the fit is unstable when the lag range is 45 lag days or more.

Fig. 3 Autocorrelation plot of the daily admissions

This result suggests that the health effect of PM_{10} exposure may be too small to estimate efficiently after 45 days in our study design. A lag range of 30 or 40 days is widely used in this issue [4, 5, 11, 12]. A shorter lag period of a week or less was also suggested by Schwartz et al. [13] for long-term trends or seasonal terms. A good practice would be to select an appropriate lag range by sensitivity analysis when DLNM is used.

One interesting result of the best-fitted model is that the relative risk estimated for low PM_{10} intensities and high lag days is relatively high as shown in the nearest corner of Fig. 1(1.1). This result may exhibit the so-called mortality displacement effect [14]. This effect decomposes the lag window into three successive periods. The first period represents the positive effect on morbidity or mortality for short lags of air

pollution exposure. The second period shows a zero or negative effect for longer lags because of the depletion of vulnerable individuals who have had their death or admission brought forward by a few days or weeks by air pollution. The last period demonstrates a positive effect when the depletion effect subsides, or when vulnerable individuals are recruited because of the net result of aging and environmental factors, such as infection, smoking, and occupation. The relatively higher risk at the end of the lag window may correspond to the last period of the mortality displacement effect.

The health effects of NO_2 and SO_2 on the daily respiratory disease morbidity are insignificant. These results are influenced by the fact that the PM_{10} intensity is much higher than the SO_2 and NO_2 levels for most days in the study period (at

Fig. 4 Autocorrelation plot of the residuals of the best-fitted model

Fig. 5 Seasonal subseries plot of the daily admissions

1794 of total 2192 days, 81.84%). The variation of the SO_2 and NO_2 intensity is also smaller as shown in Table 2. This result coincides with those of Oftedal et al. [15], who suggested that small variations in the air pollutant intensity reduce the precision of the estimated health effects.

The health effect of other air pollutants is explored in many researches, e.g., fine particulate matter ($\text{PM}_{2.5}$) [2, 9, 16]. Conclusions different from this paper are drawn. For instance, NO and NO_2 are two main air pollutants that are positively associated with respiratory diseases [17], as CO and NO_2 in Qatar [18], and benzene in Drammen, Norway [15]. However, the data of $\text{PM}_{2.5}$, NO , CO , and benzene was not collected until 2015 in China. So the data is not available and its health effect is not estimated in this study. On the other hand, those air pollutants show significant health effect only in particular local area or environment. In Chongqing, PM_{10} is more significant than NO_2 and SO_2 , as shown in our results.

Since this study focuses on estimating the exposure-lag-response association between air pollution and local respiratory disease morbidity on a daily basis, some potentially relevant factors are not investigated, including gender, age, and smoking subgroups. The subgroup distribution of the factors neglected does not vary day by day in a particular city.

In addition, some confounders, which are significant in other papers, are not included in the best-fitted model, e.g., seasonal indicator. A seasonal subseries plot of the daily admissions is shown in Fig. 5. As we can see, the means for each month are relatively close and show no obvious pattern. Chongqing is a hot city. The mean $\pm$ SD of the daily maximum temperature is 22.74 ± 9.30 °C. Meanwhile, a correlation test between the 8-day moving average of the maximum temperature and the seasonal indicator suggests high association. The Pearson's correlation coefficient is 0.828 ± 0.013 for 95% confidence interval.

The result of the Augmented Dickey-Fuller (ADF) test shows that the two time series are nonstationary and cointegrated. That is the reason why the confounder term fitted is a smooth term of the last 8-day moving average of the maximum temperature, which may be more significant than the seasonal indicator.

6 Conclusion

The modeling framework of DLNM combining with GAM is commonly used to estimate the health effects of protracted exposure events of varying intensities sustained with time. In this paper, we demonstrate the fitting and testing processes used to quantify the health risk of air pollution on respiratory disease morbidity in Chongqing City, China. With meticulous model selection and result analysis, our results suggest that the model fits the data well. There are different results and conclusions existed in other related studies. However, all the differences may come from the different environment of the research areas. The result model can be used to predict the relative risk of increased air pollution or the cumulative health effects of a history of exposure to air pollution in Chongqing within a limited lag range of 40 days.

Acknowledgments This research is supported by Young and Middle-age High-level Medical Reserved Personnel Training Project Foundation of Chongqing, China.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., Horak Jr., F., Puybonnieux-Textier, V., Quénel, P., Schneider, J., Seethaler, R., Vergnaud, J. C., & Sommer, H. (2000). Public-health impact of outdoor and traffic-related air pollution: a European assessment. *Lancet*, 356(9232), 795–801.
- Pope, C. A. (2000). Epidemiology of fine particulate air pollution and human health: biologic mechanisms and who's at risk? *Environmental Health Perspectives Supplements*, 108(suppl 4), 713–723.
- Atkinson, R. W., Anderson, H. R., Sunyer, J., et al. (2001). Acute effects of particulate air pollution on respiratory admissions—results from APHEA 2 project. *American Journal of Respiratory & Critical Care Medicine*, 164(10 Pt 1), 1860–1866.
- Gasparrini, A. (2014). Modeling exposure–lag–response associations with distributed lag non-linear models. *Statistics in Medicine*, 33(5), 881–899.
- Gasparrini, A., Armstrong, B., & Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29(21), 2224–2234.
- Gasparrini, A. (2011). Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, 43(8), 1–20.
- Nelder, J. A., & Baker, R. J. (1995). Generalized linear models. *Statistics for Biology & Health*, 2(3), 291–303.
- Trevor, H., & Tibshirani, R. (1984). Generalized additive models. *Statistical Science*, 1(42), 297–310.
- Zhou, M., He, G., Fan, M., Wang, Z., Liu, Y., Ma, J., Ma, Z., Liu, J., Liu, Y., Wang, L., & Liu, Y. (2015). Smog episodes, fine particulate pollution and mortality in China. *Environmental Research*, 136, 396–404.
- Ambient air quality standards of China (GB 3095-1996). <http://www.sps.gov.cn/page/CN/1996/GB%203095-1996.shtml>. 8 Oct 2015.
- José, F., Anabela, B., Aida, S., et al. (2011). The lag structure and the general effect of ozone exposure on pediatric respiratory morbidity. *International Journal of Environmental Research & Public Health*, 8(10), 4013–4024.
- Zhang, Y., Peng, L., Kan, H., Xu, J., Chen, R., Liu, Y., & Wang, W. (2014). Effects of meteorological factors on daily hospital admissions for asthma in adults: a time-series analysis. *PLoS One*, 9(7), e102475.
- Schwartz, J., Spix, C., Touloumi, G., et al. (1996). Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *Journal of Epidemiology & Community Health*, 50 suppl 1(2), S3–s11.
- Zanobetti, A., Wand, M. P., Schwartz, J., & Ryan, L. M. (2000). Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics*, 1(3), 279–292.
- Oftedal, B., Nafstad, P., Magnus, P., Bjørkly, S., & Skrandal, A. (2003). Traffic related air pollution and acute hospital admission for respiratory diseases in Drammen, Norway 1995-2000. *European Journal of Epidemiology*, 18(7), 671–676.
- Sinclair, A. H., Edgerton, E. S., Wyzga, R., & Tolsma, D. (2010). A two-time-period comparison of the effects of ambient air pollution on outpatient visits for acute respiratory illnesses. *Journal of the Air & Waste Management Association*, 60(2), 163–175.
- Wang, K. Y., & Chau, T. T. (2013). An association between air pollution and daily outpatient visits for respiratory disease in a heavy industry area. *PLoS One*, 8(10), e75220.
- Bener, A., Dogan, M., & Shanks, N. J. (2009). The impact of air pollution on hospital admission for respiratory and cardiovascular diseases in an oil-rich country. *East African Journal of Public Health*, 41(3), 124–127.