Supplemental Material

The burden of air pollution on years of life lost in Beijing, China, 2004-08: retrospective analysis of daily deaths

Yuming Guo *Research Fellow*[1], Shanshan Li *PhD Candidate*[1], Zhaoxing Tian *Doctor of Emergency Medicine*[2], Xiaochuan Pan *Professor of Environmental Health*[3], Jinliang Zhang *Professor of Environmental Health*[4], Gail Williams *Professor of International Health and Statistics*[1]

1. Department of Epidemiology and Biostatistics, School of Population Health, University of Queensland, Brisbane, Australia.

2. Emergency Department of Peking University Third Hospital, Beijing, China.

3. Department of Occupational and Environmental Health, Peking University School of Public Health, Beijing, PR China.

4. State Key Laboratory of Environmental Criteria and Risk Assessment & Environmental Standards Institute, Chinese Research Academy of Environmental Sciences, Beijing, China.

Corresponding authors

Dr Yuming Guo, Department of Epidemiology and Biostatistics, School of Population Health, University of Queensland, Brisbane, Queensland 4006, Australia.

Email: y.guo1@uq.edu.au.

Dr Zhaoxing Tian, Emergency department of Peking University Third Hospital, Beijing 100191, China.

Email: tjtzx@126.com.

Table of contents

Example for calculating years of life lost using R software

We used a simulated data as an example to show how to calculate years of lost for each death, and how to sum daily years of life lost for each group (e.g., gender and age groups). Please be aware that this data (BMJ.example.csv) is not the real dataset used in our study.

We caculate years of life (yll) for each death by matching age and sex to the life tables (Supplemental Material Table S1). To make it easy to understand, we calculate years of life lost for male and female separately.

```
## read the simulated dataset
data<-read.csv("BMJ.example.csv")  ## Please download the data from BMJ website
data$deathday<-as.Date(data$deathday,"%d/%m/%Y")

 ## First for male years of life lost
data1<-subset(data,data$sex%in%"male")
data1$yll[data1$age<1]<-71.1
data1$yll[data1$age>=1&data1$age<5]<-71.4
data1$yll[data1$age>=5&data1$age<10]<-67.5
data1$yll[data1$age>=10&data1$age<15]<-62.9
data1$yll[data1$age>=15&data1$age<20]<-58
data1$yll[data1$age>=20&data1$age<25]<-53.3
data1$yll[data1$age>=25&data1$age<30]<-48.6
data1$yll[data1$age>=30&data1$age<35]<-43.9
data1$yll[data1$age>=35&data1$age<40]<-39.2
data1$yll[data1$age>=40&data1$age<45]<-34.6
data1$yll[data1$age>=45&data1$age<50]<-30
data1$yll[data1$age>=50&data1$age<55]<-25.6
data1$yll[data1$age>=55&data1$age<60]<-21.5
data1$yll[data1$age>=60&data1$age<65]<-17.6
data1$yll[data1$age>=65&data1$age<70]<-14.1
data1$yll[data1$age>=70&data1$age<75]<-11
data1$yll[data1$age>=75&data1$age<80]<-8.3
data1$yll[data1$age>=80&data1$age<85]<-6.1
data1$yll[data1$age>=85&data1$age<90]<-4.4
data1$yll[data1$age>=90&data1$age<95]<-3.2
data1$yll[data1$age>=95&data1$age<100]<-2.3
data1$yll[data1$age>=100]<-1.8

## Second for female years of life lost
data2<-subset(data,data$sex%in%"female")
```

```
data2$yll[data2$age<1]<-74.2
data2$yll[data2$age>=1&data2$age<5]<-75.3
data2$yll[data2$age>=5&data2$age<10]<-71.6
data2$yll[data2$age>=10&data2$age<15]<-66.8
data2$yll[data2$age>=15&data2$age<20]<-61.9
data2$yll[data2$age>=20&data2$age<25]<-57
data2$yll[data2$age>=25&data2$age<30]<-52.1
data2$yll[data2$age>=30&data2$age<35]<-47.3
data2$yll[data2$age>=35&data2$age<40]<-42.5
data2$yll[data2$age>=40&data2$age<45]<-37.8
data2$yll[data2$age>=45&data2$age<50]<-33.1
data2$yll[data2$age>=50&data2$age<55]<-28.6
data2$yll[data2$age>=55&data2$age<60]<-24.2
data2$yll[data2$age>=60&data2$age<65]<-19.9
data2$yll[data2$age>=65&data2$age<70]<-15.9
data2$yll[data2$age>=70&data2$age<75]<-12.3
data2$yll[data2$age>=75&data2$age<80]<-9.2
data2$yll[data2$age>=80&data2$age<85]<-6.7
data2$yll[data2$age>=85&data2$age<90]<-4.7
data2$yll[data2$age>=90&data2$age<95]<-3.3
data2$yll[data2$age>=95&data2$age<100]<-2.4
data2$yll[data2$age>=100]<-1.8

## combine data for female and male
data<-rbind(data1,data2)

################# Caculate daily series of years of life lost
## For all non-accidental deaths
yll.all<-as.numeric(tapply(data$yll,data$deathday,sum))
## For male
data.male<-subset(data,data$sex%in%"male")
yll.male<-as.numeric(tapply(data.male$yll,data.male$deathday,sum))
## For female
data.female<-subset(data,data$sex%in%"female")
yll.female<-as.numeric(tapply(data.female$yll,data.female$deathday,sum))
## For age 0-65
data065<-subset(data,data$age<=65)
yll.065<-as.numeric(tapply(data065$yll,data065$deathday,sum))
## For age >65
data65<-subset(data,data$age>65)
yll.65over<-as.numeric(tapply(data65$yll,data65$deathday,sum))
###### Put all groups into the same dataset
daily.yll<-data.frame(date=unique(data$deathday),
                      yll.all,yll.male,yll.female,yll.065,yll.65over)
```

The “daily.yll” dataset includes daily years of life lost for each group. Then we can link this dataset to dataset including daily air pollutants and weather conditions.

Sensitivity analysis

To check the robustness of our models, we performed several sensitivity analyses for associations between air pollution and years of life lost.

The life expectancy given in life tables was fixed for age and sex. However, at an individual level, life expectancy will be variable, around this average, fixed value. To account for this uncertainty, we created 10,000 datasets with simulated random life expectancies (created by sampling a normal distribution with mean equal to the life table value) for each death (refer supplemental material, Simulation for random life expectancy, for details). The mean effect estimates using a random life expectancy were similar to those based on a fixed life expectancy, but the confidence intervals were slightly wider (results not shown). The relationships between pollutants and years of life lost remained linear when we used 5 *df* NCS for air pollutants (Supplemental Material, Figure S7). The effect estimates of air pollutants on years of life lost did not change when we varied *df* and lag pattern for temperature, relative humidity, and air pressure, and used different *df* for time to control for long-term trend and season (Supplemental Material, Figure S8). When we removed temperature or other weather conditions from the models, the effect estimates remained similar (Supplemental Material, Figures S9 and S10). The autocorrelation function test shows the residuals were independent over time (Supplemental Material, Figure S11).

Simulation for random life expectancy

In our main analyses, we used fixed life expectancy (mean between the years 2000 and 2009) for each death by age and sex (Table S1). However, life expectancy for each death is uncertain. Some people may have a longer life expectancy than the fixed, while others may have a shorter life expectancy than the fixed. To make sure our results are robust based on the fixed life expectancy, we simulated random life expectancy for each death as following steps.

Step I: We simulated a normal distribution of life expectancy based on fixed life expectancy matching by sex and age. The fixed life expectancy is used as the mean of simulated life expectancy. The lowest tail of the simulated life expectancy is 0. For example the distribution of simulated life expectancy for male at age 45 and 90 is showing in Figure S1.

Step II: For each death, the life expectancy was randomly sampled from the simulated life expectancy by matching sex and age. The daily total years of life lost were calculated by summing the years of life lost for all deaths on the same day.

Step III: Analyse the relationships between air pollutants and years of life lost using data from step II.

We repeated step II and step III for 10,000 times with simulated random life expectancies. We got the means and confidence intervals for the estimated effects of air pollutants on years of life lost. All the results using a simulated random life expectancy were similar to those based on using a fixed life expectancy, but the confidence intervals were slightly wider.

Table S1: Life expectancy for Chinese population

(http://www.who.int/whosis/database/life_tables/life_tables.cfm).

Age	Life expectancy for Male (year)			Life expectancy for Female (year)		
	2000	2009	Mean	2000	2009	Mean
0	69.9	72.2	71.05	72.6	75.8	74.2
1	70.6	72.2	71.4	74.2	76.3	75.25
5	67.1	68.4	67.75	70.7	72.5	71.6
10	62.3	63.5	62.9	65.9	67.6	66.75
15	57.4	58.6	58	61.0	62.7	61.85
20	52.7	53.8	53.25	56.1	57.8	56.95
25	48.1	49.1	48.6	51.3	52.9	52.1
30	43.4	44.4	43.9	46.5	48.1	47.3
35	38.7	39.7	39.2	41.7	43.2	42.45
40	34.1	35.0	34.55	37.0	38.5	37.75
45	29.6	30.4	30	32.4	33.8	33.1
50	25.2	26.0	25.6	27.9	29.2	28.55
55	21.1	21.8	21.45	23.5	24.8	24.15
60	17.2	17.9	17.55	19.3	20.5	19.9
65	13.8	14.3	14.05	15.4	16.4	15.9
70	10.7	11.2	10.95	11.9	12.7	12.3
75	8.1	8.4	8.25	8.8	9.5	9.15
80	5.9	6.2	6.05	6.4	6.9	6.65
85	4.3	4.5	4.4	4.6	4.8	4.7
90	3.1	3.2	3.15	3.2	3.4	3.3
95	2.3	2.3	2.3	2.3	2.4	2.35
100	1.7	1.8	1.75	1.8	1.8	1.8

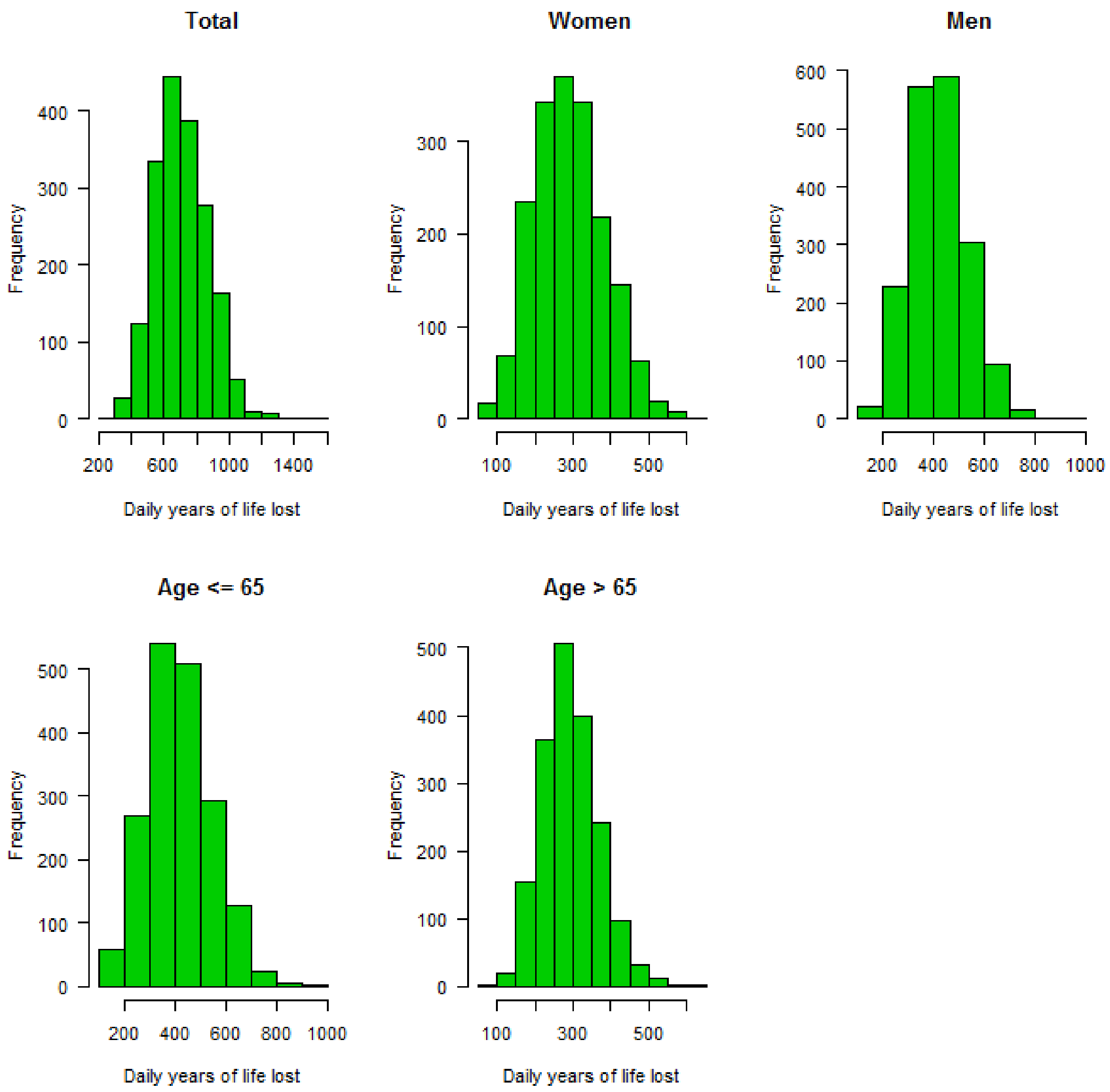

Figure S1: Histogram of daily years of life lost in Beijing, China during 2004–2008.

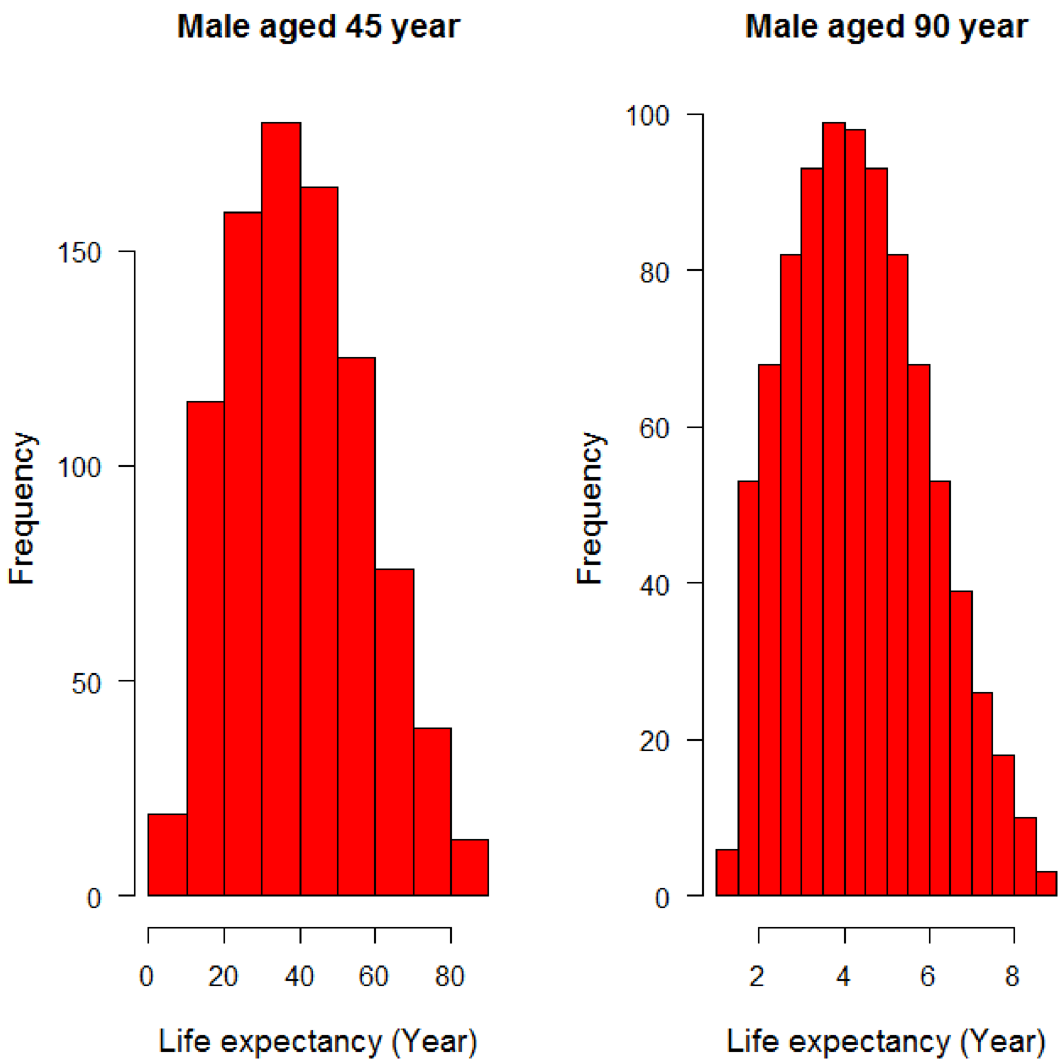

Figure S2: the distribution of simulated life expectancy for male at age 45 and 90.

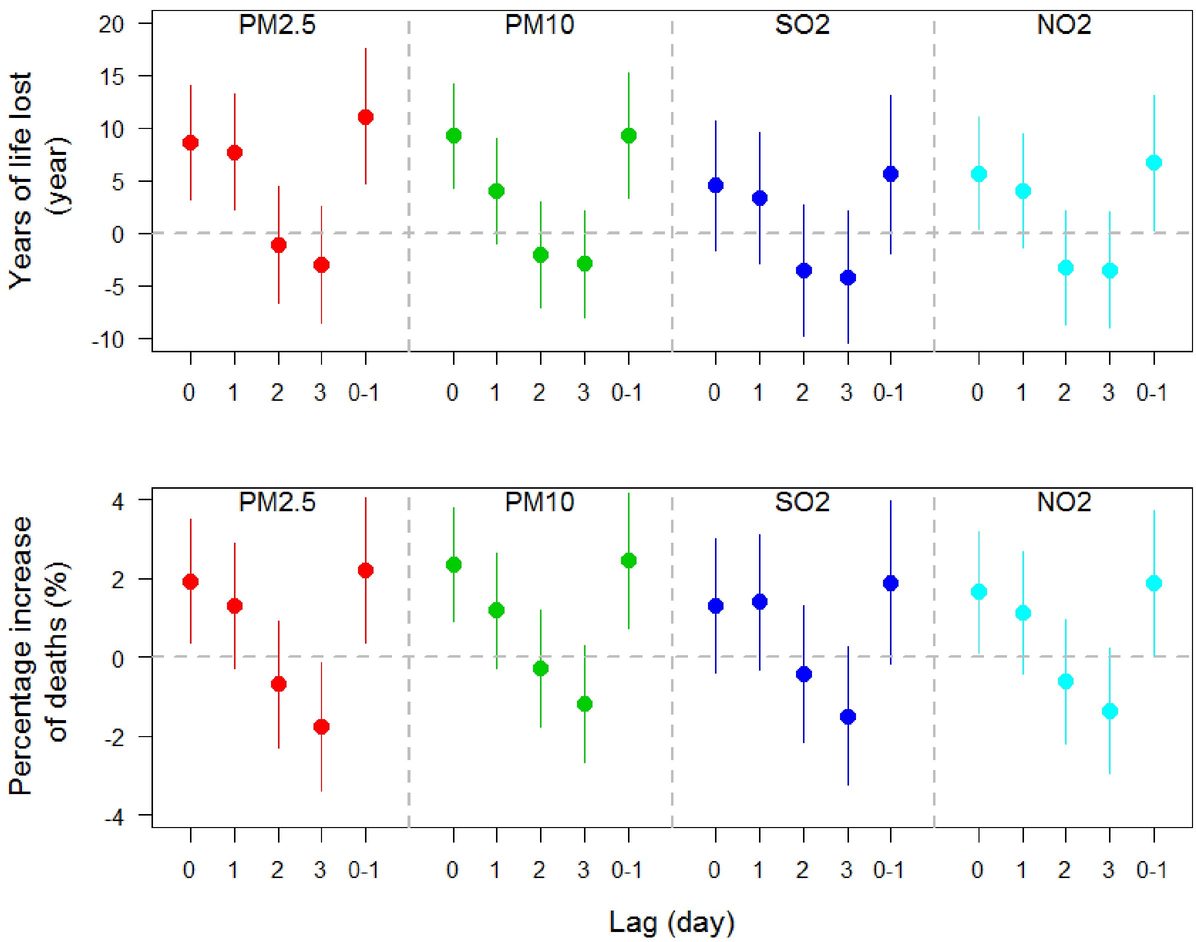

Figure S3: The relationship between a IQR increase in air pollutant and years of life lost (top) and percentage increase of deaths (bottom) for women using single pollutant models at different lag days during 2004–2008, while controlling for seasonality, day of the week, temperature, relative humidity, and air pressure. The IQRs are 94 μg/m^3 for $PM_{2.5}$, 106 μg/m^3 PM_{10}, 49 μg/m^3 SO_2, and 30 μg/m^3 NO_2.

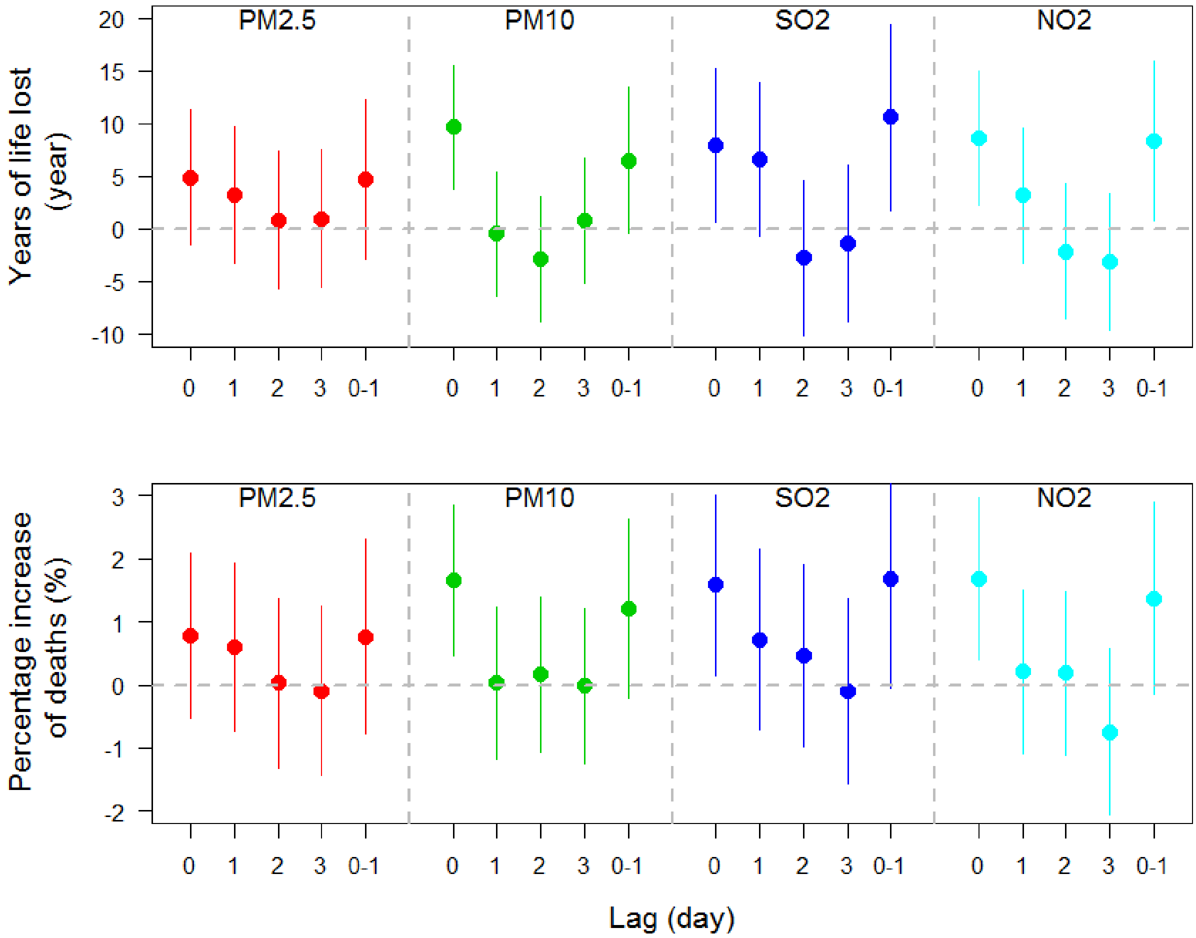

Figure S4: The relationship between a IQR increase in air pollutant and years of life lost (top) and percentage increase of deaths (bottom) for men using single pollutant models at different lag days during 2004–2008, while controlling for seasonality, day of the week, temperature, relative humidity, and air pressure. The IQRs are 94 μg/m^3 for $PM_{2.5}$, 106 μg/m^3 PM_{10}, 49 μg/m^3 SO_2, and 30 μg/m^3 NO_2.

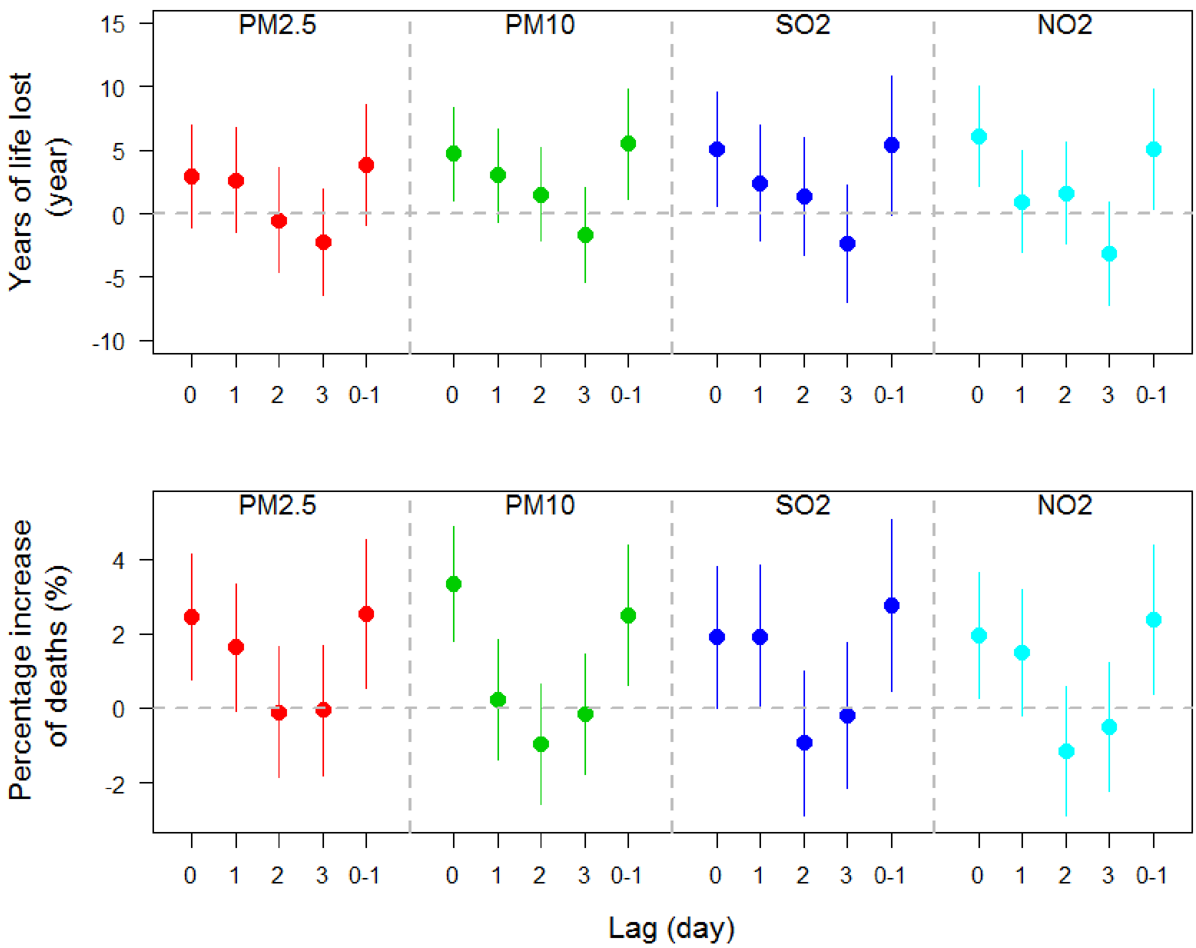

Figure S5: The relationship between a IQR increase in air pollutant and years of life lost (top) and percentage increase of deaths (bottom) for people aged <=65 years using single pollutant models at different lag days during 2004–2008, while controlling for seasonality, day of the week, temperature, relative humidity, and air pressure. The IQRs are 94 μg/m^3 for $PM_{2.5}$, 106 μg/m^3 PM_{10}, 49 μg/m^3 SO_2, and 30 μg/m^3 NO_2.

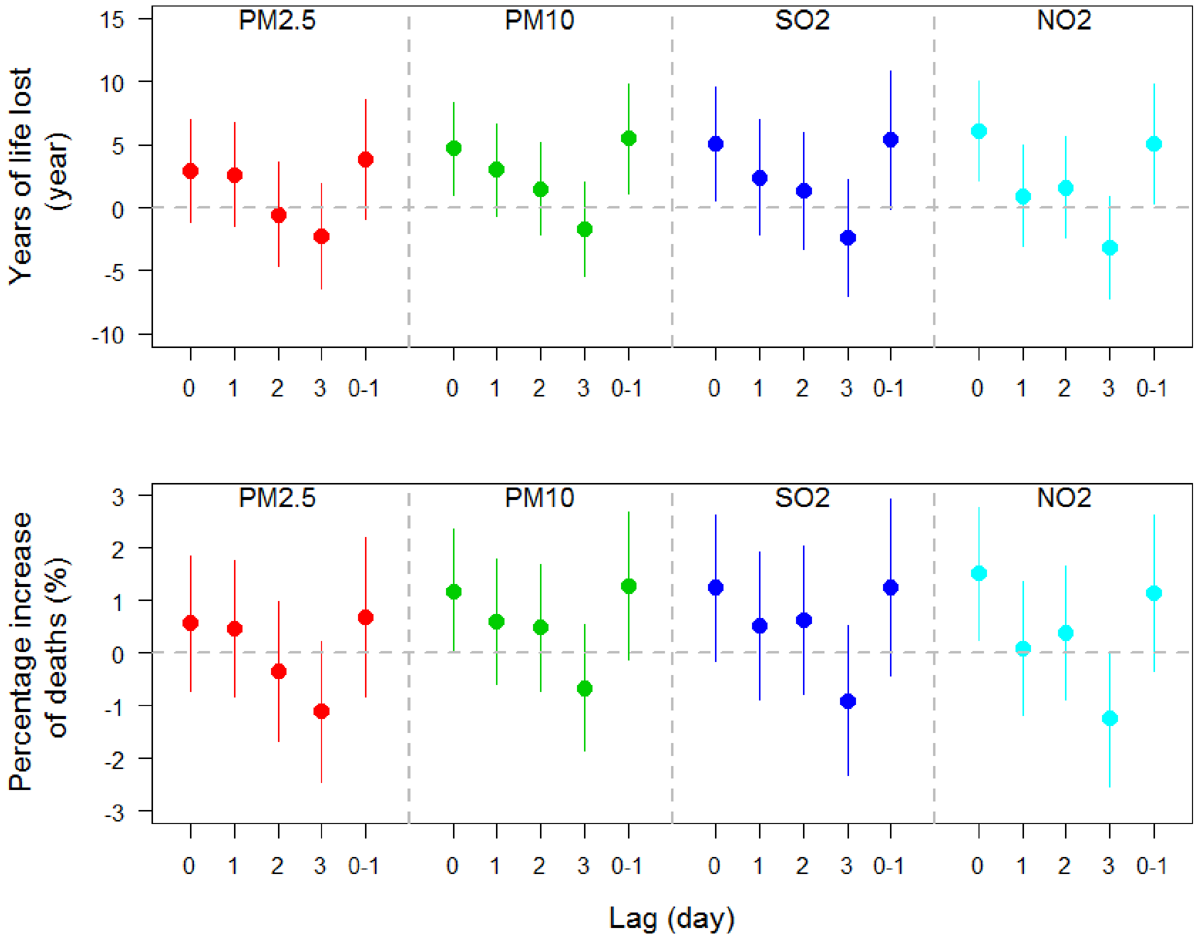

Figure S6: The relationship between a IQR increase in air pollutant and years of life lost (top) and percentage increase of deaths (bottom) for people aged > 65 years using single pollutant models at different lag days during 2004–2008, while controlling for seasonality, day of the week, temperature, relative humidity, and air pressure. The IQRs are 94 μg/m^3 for $PM_{2.5}$, 106 μg/m^3 PM_{10}, 49 μg/m^3 SO_2, and 30 μg/m^3 NO_2.

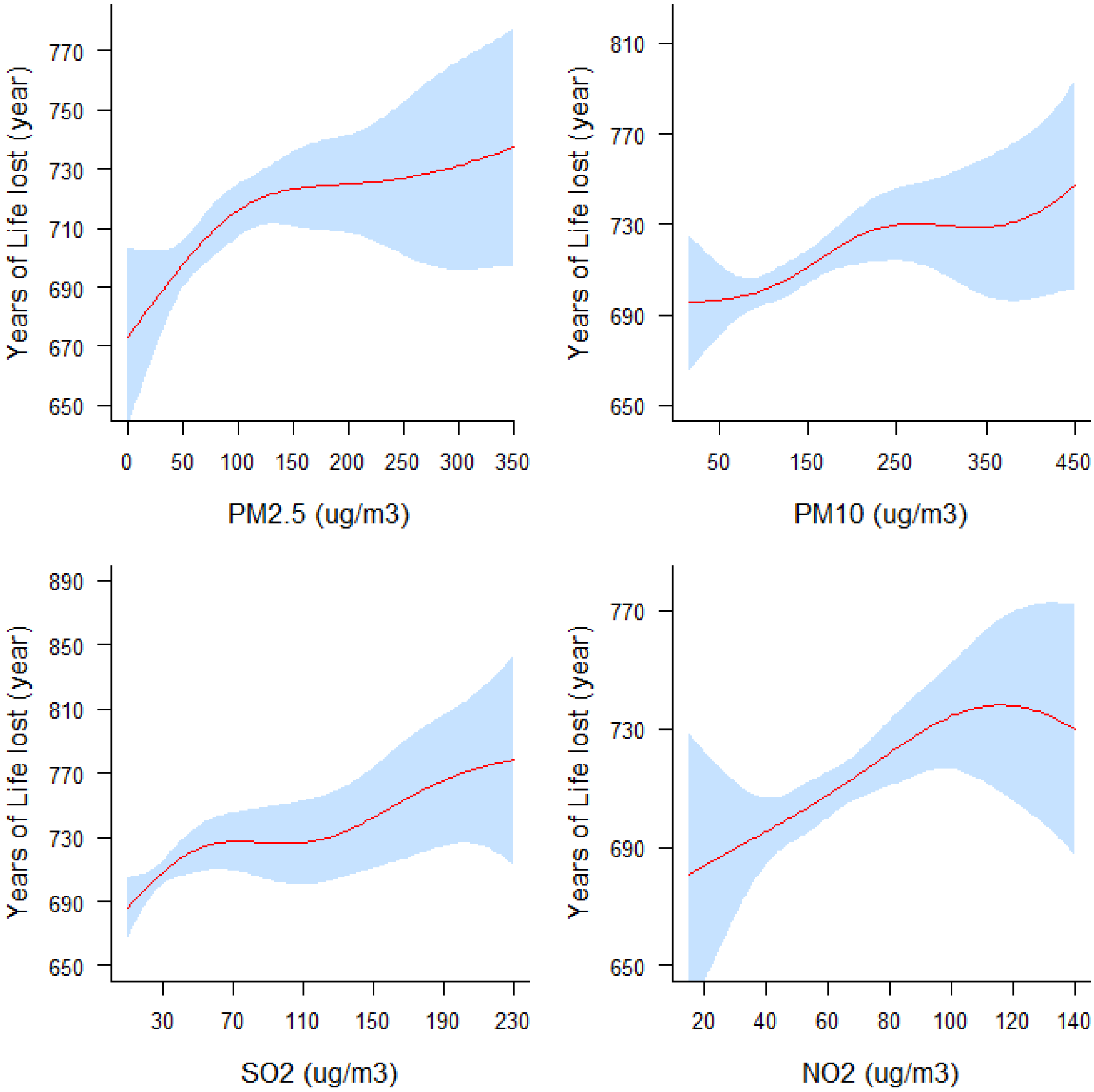

Figure S7: The relationship between air pollutants (lag 0–1) and years of life lost in Beijing China, during 2004–2008. A natural cubic spline with 5 degrees of freedom for air pollutants was included in the single pollutant models, while controlling for seasonality, day of the week, temperature, relative humidity, and air pressure.

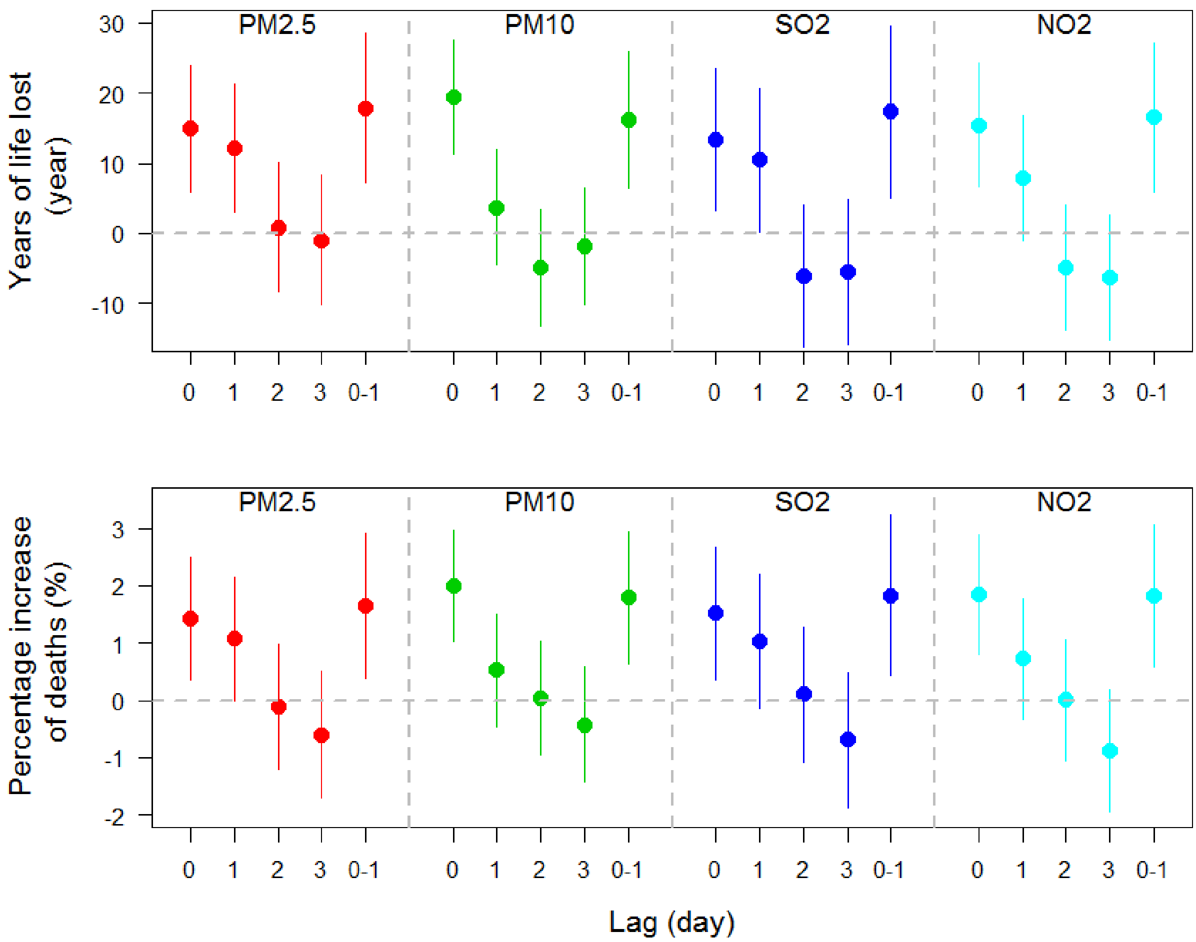

Figure S8: The relationship between a IQR increase in air pollutant and years of life lost (top) and percentage increase of deaths (bottom) for non-accidental deaths using single pollutant models at different lag days during 2004–2008, while controlling for seasonality, day of the week, temperature, relative humidity, and air pressure. A natural cubic spline with 10 degrees of freedom per year for time was used. The IQRs are 94 μg/m^3 for $PM_{2.5}$, 106 μg/m^3 PM_{10}, 49 μg/m^3 SO_2, and 30 μg/m^3 NO_2.

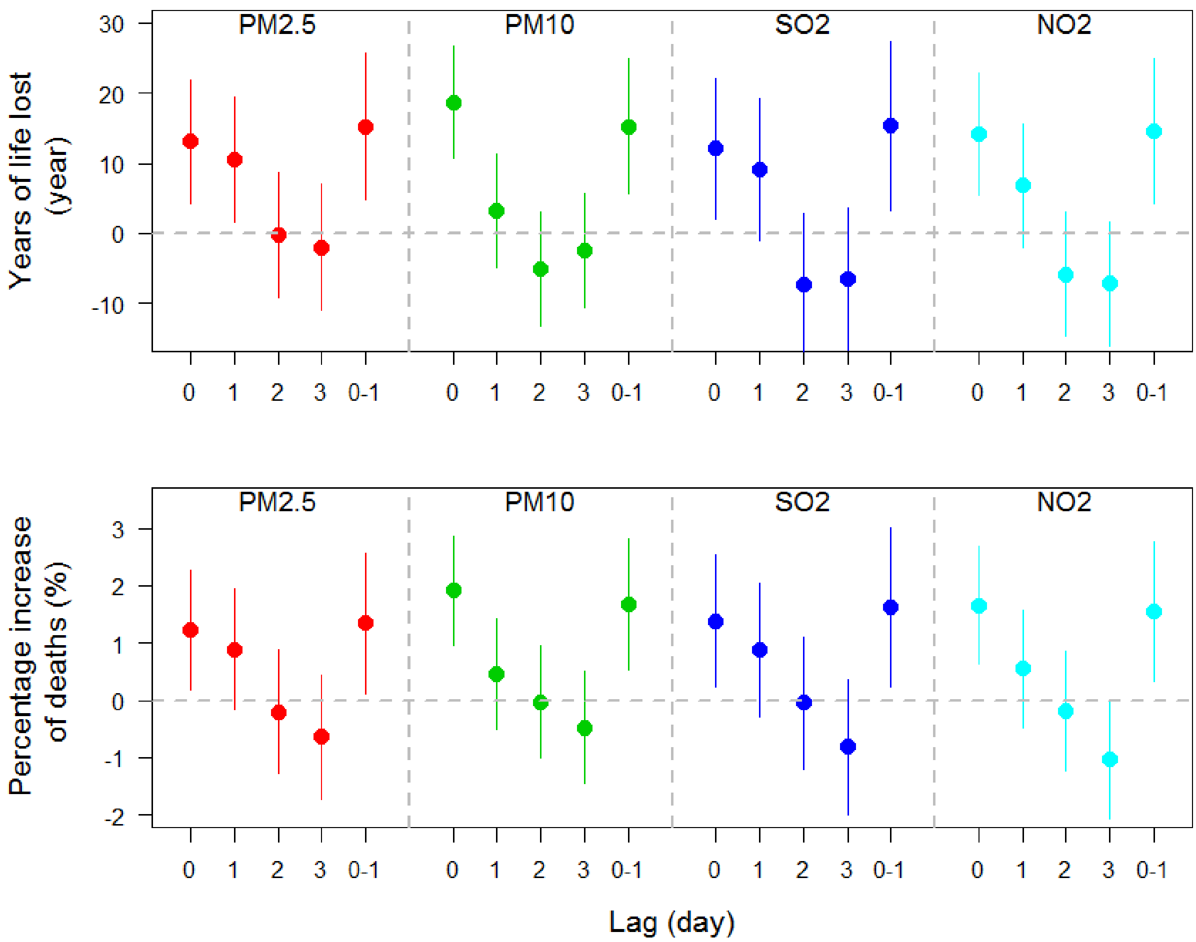

Figure S9: The relationship between a IQR increase in air pollutant and years of life lost (top) and percentage increase of deaths (bottom) for non-accidental deaths using single pollutant models at different lag days during 2004–2008, while controlling for seasonality, day of the week, relative humidity, and air pressure. The IQRs are 94 μg/m^3 for $PM_{2.5}$, 106 μg/m^3 PM_{10}, 49 μg/m^3 SO_2, and 30 μg/m^3 NO_2.

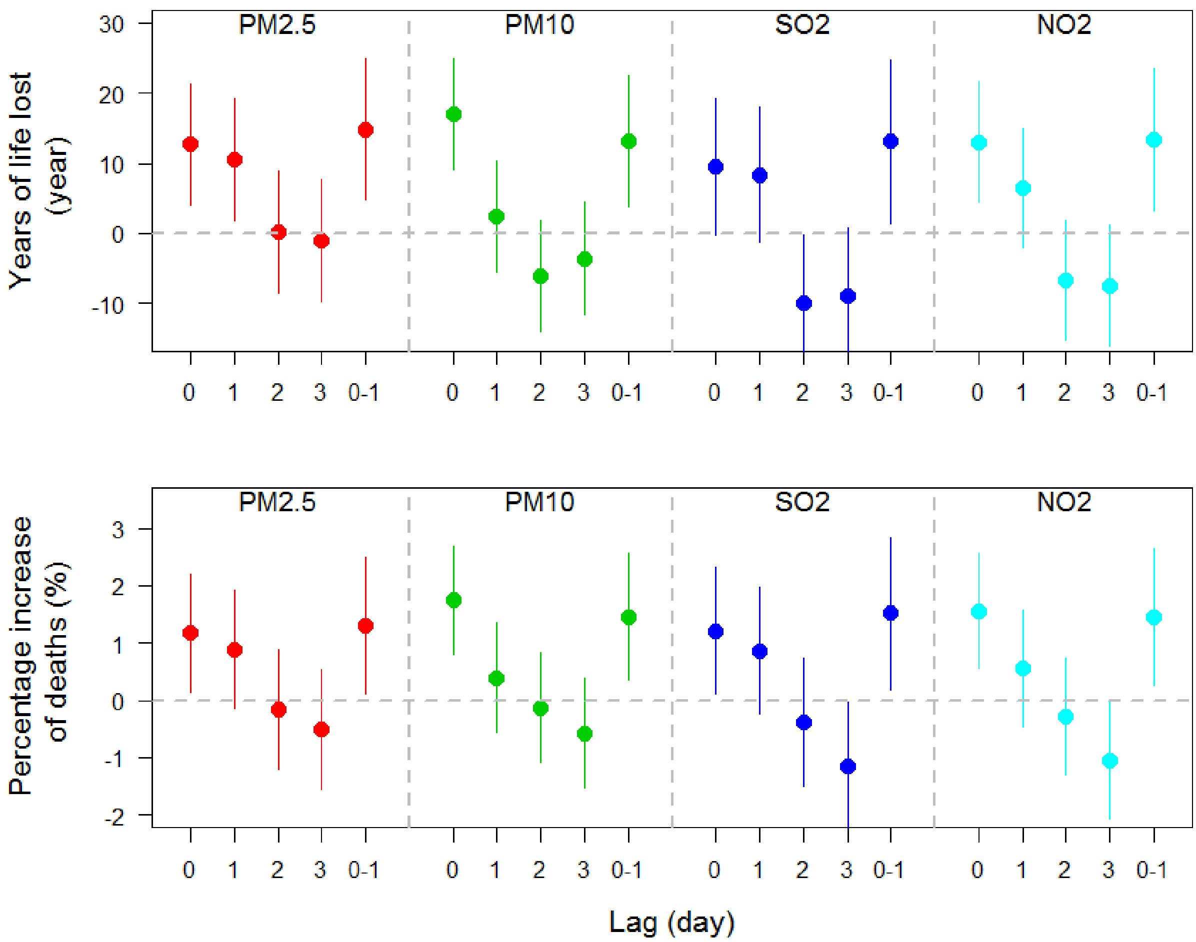

Figure S10: The relationship between a IQR increase in air pollutant and years of life lost (top) and percentage increase of deaths (bottom) for non-accidental deaths using single pollutant models at different lag days during 2004–2008, while controlling for seasonality, day of the week. The IQRs are 94 μg/m^3 for $PM_{2.5}$, 106 μg/m^3 PM_{10}, 49 μg/m^3 SO_2, and 30 μg/m^3 NO_2.

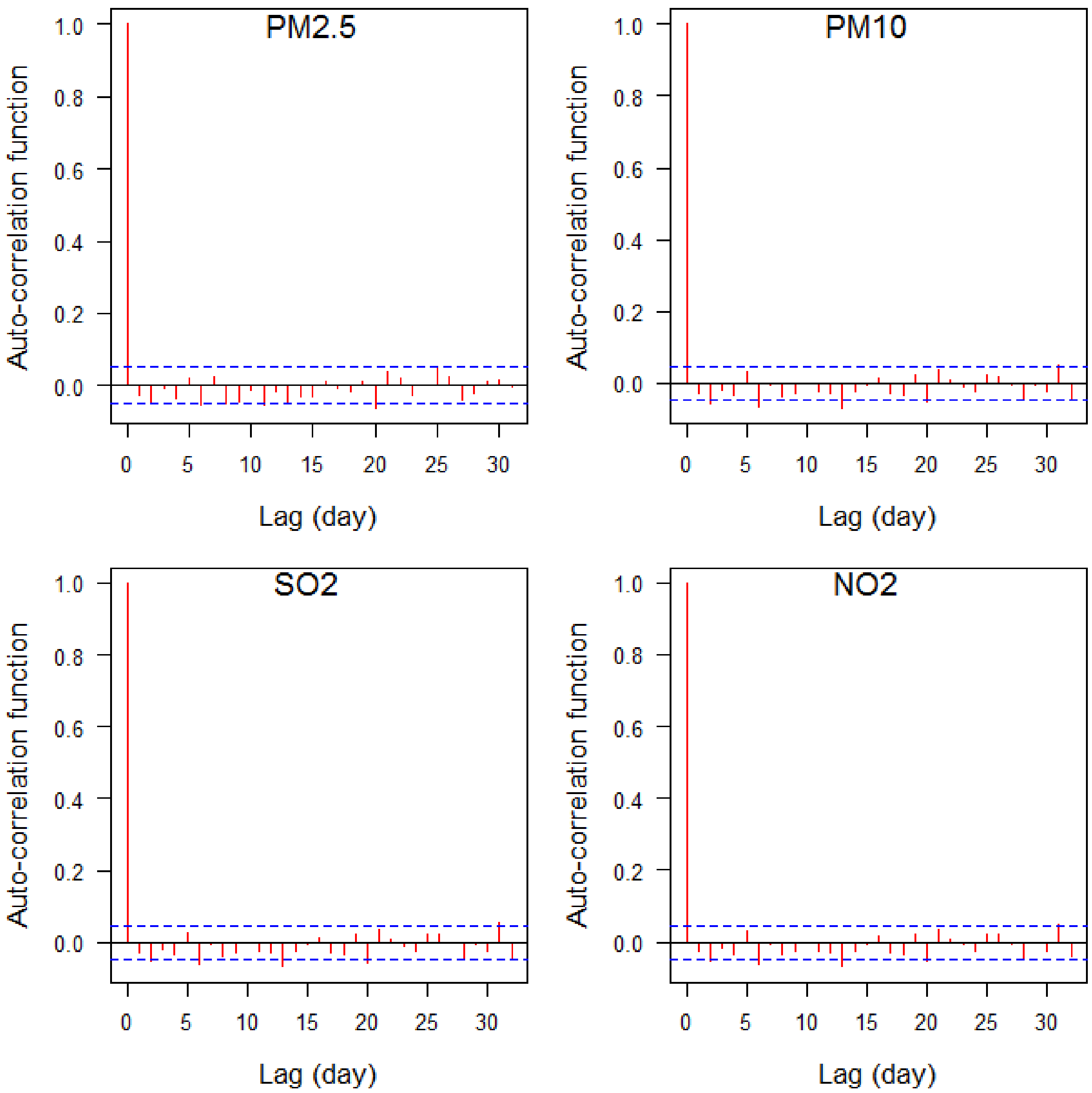

Figure S11: Auto-correlation function for residuals of single pollutant models.