Distribution Agreement

Effect modification by environmental quality index

on the short-term association between $PM_{2.5}$ and mortality

By

Tong Wang

MSPH

Emory University

Rollins School of Public Health

Department of Biostatistics

____ [Chair's signature]

Howard H. Chang

____ [Member's signature]

James L. Crooks

Effect modification by environmental quality index

on the short-term association between $PM_{2.5}$ and mortality

By

Tong Wang

B.S., Wuhan University, 2009

MPH, Emory University

Rollins School of Public Health

2015

Advisor: Howard H. Chang

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health

in Department of Biostatistics

2015

Abstract

Effect modification by environmental quality index on the short-term association between $PM_{2.5}$ and mortality

By Tong Wang

Abstract:

Here we used the data obtained from US Environmental Protection Agency to examine association between $PM_{2.5}$ and total non-accidental death risk and to evaluate the heterogeneity in $PM_{2.5}$ mortality rate across different counties. We also evaluated whether county-level environmental quality index modifies the association between $PM_{2.5}$ and total non-accidental death risk. This study included data of 433 US counties. Bayesian meta-regression was used to combine relative risks from 433 US counties from a national multisite time-series analysis. Among the five environment quality indexes to represent different domains (air, built environment, land, water, sociodemographic) considered, air quality and built environment were found to be significant effect modifiers on $PM_{2.5}$ health risks. The effect modification were robust after adjustment by other environmental quality indexes.

Effect modification by environmental quality index

on the short-term association between $PM_{2.5}$ and mortality

By

Tong Wang

B.S., Wuhan University, 2009

MPH, Emory University

Rollins School of Public Health

2015

Advisor: Howard H. Chang

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health

in Department of Biostatistics

2015

Table of Contents

Chapter I 1

Introduction 1

Chapter II 11

Literature Review 11

Chapter III 12

Methods 12

Chapter IV 15

Results 15

Chapter V 17

Discussion 17

Reference 18

Appendix 20

Chapter I

Introduction

Particulate matter (PM) and human health

Particulate matter (PM) is defined as the minute mixture of solid particles or liquid droplets that are suspended in the air.[1] Based on their penetration capacity into the lungs, Environmental Protection Agency(EPA) has been categorizing particles mainly into two categories: coarse particulate matter (PM_{10}) with an aerodynamic diameter of less than 10 μm and fine particulate matter ($PM_{2.5}$) with an aerodynamic diameter of 2.5 μm[2]. Particulate matters are primarily generated from sources such as road dust, agriculture dust, river beds, construction sites, mining operations, and similar human activities[3]. PM is a key indicator of air pollution. The chemical constituents of PM include nitrates; sulfates; elemental and organic carbon; organic compounds; biological compounds; and metals such as iron, copper, nickel, zinc, and vanadium.[4] Since PM can be suspended over a long time and travel over long distances in the atmosphere, it has the potential to cause or exacerbate a variety of diseases.[5] Several studies have reported association between exposure to PM and total morality, as well as morality due to cardiovascular and respiratory diseases.[1,6] More than two million deaths are estimated to occur globally each year as a direct consequence of air pollution due to the respiratory disease [7]. Among these deaths, around 2.1 millions are caused by $PM_{2.5}$. [7–8] Because of the potential adverse health effects of $PM_{2.5}$ and its associated pollutants, detailed knowledge of the association between PM and human health is of primary importance.[5]

Effect modification on $PM_{2.5}$ mortality risk

Multisite population-based epidemiological studies have observed significant heterogeneity in the regional difference in the $PM_{2.5}$ mortality risk estimates. [9]The observed regional differences in PM mortality risk estimates have often been attributed to a variety of factors, including geographic variability in particle composition, spatial heterogeneity of constituents, and differences between cities in the distribution of the population potentially at greatest risk of an air pollutant-related health effect.[10]

A better understanding of the effect modification on mortality risk can help us understand the regional difference in heterogeneity and make corresponding policy to reduce the mortality rate.

Environmental quality index(EQI)

To help us better understand the relationship between environment conditions and human health, an environmental quality index(EQI) for all counties in the US was recently developed. By incorporating a variety of environmental and population data, five environmental domains were constructed(air, water, land, built environment and sociodemographic). The unit of analysis for EQI development was at each U.S. county and the indexes were developed using data between 2002 to 2005. Here we briefly describe the data sources and variable construction of each EQI.

Data sources

Daily concentrations of six criteria air pollutants from the Air Quality System(AQS) and county-level hazardous air pollutants(HAP) concentrations from the NationalScale Air Toxics Assessment(NATA) were included and calculated in the air domain.

Eleven variables from the U.S. Census and data on area-level crime environment from the Federal Bureau of Investigation (FBI) Uniform Crime Report (UCR) were employed in the sociodemographic domain.

Water domain included five data sources: Watershed Assessment, Tracking & Environmental Results (WATERS) Program Database, Estimates of Water Use in the U.S. , National Atmospheric Deposition Program (NDAP), Drought Monitor Network, and National Contaminant Occurrence Database (NCOD). Data on water impairment, water contamination, recreational water quality, quality of the water used for domestic, possible drought status conditions, chemical contamination of water supplies and concentration of nine chemicals in precipitation, calcuim, magnesium, potassium, sodium, ammonium, nitrate, chloride, sulfate, and mercury were collected and assessed from the above sources.

Land domain comprised five sources: the 2002 National Pesticide Use Database, the 2002 Census of Agriculture, the National Priority Site data, the National Geochemical Survey and the EPA Radon Zone Map. Information on the agricultural environment, herbicide, insecticide, and fungicide use, the natural geochemistry and soil contamination, large industrial facilities and potential for elevated indoor radon levels were obtained and assessed from these sources.

The built environment domain employed five data sources: the Topographically Integrated Geo- coding Encoding Reference (TIGER), the Fatality Analysis Reporting System (FARS) data, yearly report of National Highway Traffic Safety administration, Housing and Urban Development data and data collected by Dun and Bradstreet. Data on housing environment, high way safety, the proportions of highway and primary roads, business and service environment were constructed and assessed from these resources.

Variable constructions

Air domain

Daily concentrations of six criteria air pollutants from the Air Quality System (AQS) were temporally averaged to get annual mean concentrations for each monitor location from 2000 to 2005. To estimate annual concentrations at each county center point, the annual means were temporally and spatially kriged. These kriged values were then averaged for the six-year study period.

County-level hazardous air pollutants (HAP) concentrations from the NationalScale Air Toxics Assessment(NATA) were retrieved. Emission estimates for each variable were averaged to get a composite emissions estimate across the study period. Air domain variables were checked for normality and log-transformed if necessary.

Sociodemographic domain

Eleven variables from the United States Census were included in the sociodemographic domain of the EQI. When the sociodemographic domain was constructed, positive variables were reverse-coded to ensure that a higher amount of the sociodemographic domain represented adverse environmental conditions.

The area-level crime environment was represented using the data from Federal Bureau of Investigation (FBI) Uniform Crime Reports (UCR). Crime data were spatially and temporally kriged to estimate counties with no reported crime. Kriging employed a double exponential covariance structure for the spatial covariance; one represented short-range variability and the other long-range variability. The covariance model was fit to experimental covariance values with a least squares method and demonstrated sufficient fit. The crime variable was log-transformed to be included in the EQI.

Water domain

Using the WATERS database and joining the data in GIS software with measures of stream length in the Reach Address Database, a cumulative measure of percent of water impairment for agricultural, drinking, recreational wildlife and industrial use was used to represent overall water quality in the county.

The number of National Pollutant Discharge Elimination System (NPDES) permits in a county was used as a proxy for general water contamination. Three composite variables were included in the EQI: a composite for number of sewage permits, a composite for industrial permits, and one for stormwater permits, all per 1000 km of stream length per county.

Quality of recreational water was assessed using the WATERS database , from which three variables for number of days of beach closure were created - for any event, for contamination events, and for rain events in a county.

Domestic water quality data was obtained from the Estimates of Water Use in the U.S. database as a proxy for the quality of the water used for domestic needs. Two variables were therefore included in the EQI: the percent of population on self-supplied water supplies and the percent of those on public water supplies which are on surface waters.

The NDAP dataset provides measures for the concentration of nine atmospheric chemicals in precipitation, calcuim, magnesium, potassium, sodium, ammonium, nitrate, chloride, sulfate, and mercury. Annual summary data for each year from each monitoring site were spatially kriged, using an exponential covariance structure, to achieve national coverage and county level estimates. The annual estimates for each pollutant were then averaged for the full study period. The data for all pollutants, except sulfate, were log-transformed.

The Drought Monitor dataset offers raster data on six possible drought status conditions for the entire U.S. on a weekly basis. To estimate the percentage of the county in each drought status condition the data were spatially aggregated to the county level. From this data we used the percentage of the county in extreme drought (D3-D4) in the EQI.

Chemical contamination of water supplies was obtained from the NCOD dataset which provides data on 69 contaminants provided by public water supplies throughout the country for the period from 1998–2005. Data for all samples in a county for each contaminant were averaged over the time period and log-transformed. With the assumption that lack of measurement for an area indicated low concern for that particular contaminant, missing values were set to zero.

Land domain

In total, eight variables representing agriculture were constructed using data from 2002 Census of Agriculture and county-level percentages (acres applied per county total acreage) were calculated and then log-transformed.

Herbicide, insecticide, and fungicide use for each county and state pesticide use data were estimated using data from the 2002 Census of Agriculture and the 2002 National Pesticide Use Dataset, respectively. All pesticide variables were log-transformed.

The natural geochemistry and soil contamination was estimated using data from the National Geochemical Survey (NGS). These data, collected for stream sediments, soils, and other media, were combined at the county level to estimate the mean values of 13 geochemical contaminants available and were log-transformed.

Large industrial facilities represent sources for pollutants released into the environment. The National Priority List data from the EPA provided information on facilities for the U.S. Because many counties had at least one, but no counties had all six of the facility types present, a composite facilities data variable was constructed by summing the count of any one of the six facilities types (brownfield sites, superfund sites, toxic release inventory sites, pesticide producing lo- cation sites, large quantity generator sites, and treatment, storage and disposal sites) across the counties. The facilities rate variable was assessed for normality and log-transformed.

Finally, the potential for elevated indoor radon levels was represented using county score from the EPA Radon Zone map.

Built environment domain

The built environment domain employed five data sources: the Topographically Integrated Geo- coding Encoding Reference (TIGER), the Fatality Analysis Reporting System (FARS) data, yearly report of National Highway Traffic Safety administration, Housing and Urban Development data and data collected by Dun and Bradstreet. Data on housing environment, high way safety, the proportions of highway and primary roads, business and service environment were constructed and assessed from these resources

Housing environment was represented by two variables available from the HUD data source, low-rent and section-eight, which were summed to get the count of any low-rent or section-eight housing in each county; the subsidized housing rate was constructed and log-transformed from this count.

Highway safety was represented using a traffic fatality variable. Rates of count of fatal crashes per county were constructed and log distributed. The percent of county residents who use public transportation was the only U.S. Census variable and was log-transformed.

The proportions of each county that were served by highways, secondary roads and primary roads were also included.

Nine business environment rate variables were constructed using data from Duns and Bradstreet by dividing the county-level count of a business type by the county- level population count. All variables except the negative food environment were log-transformed. When the built domain was constructed, positive variables were reverse-coded to ensure that a higher amount of these service variables represent adverse environmental conditions.[11]

All of these five domains have a negative valence, meaning that the higher a specific domain score, the poorer environment quality.

Problem statement

There has been considerable research on the associations between $PM_{2.5}$ levels and mortality risk and human disease. There is increasing interest in research focusing on identifying effect modifiers that can contribute to spatial effect heterogeneity. To have a comprehensive understanding on the effect modifiers, we needs to identify whether environmental quality index modifies the association between $PM_{2.5}$ and mortality risk.

Purpose statement

In this study we aim to examine the association between $PM_{2.5}$ and total non-accidental death risk, as well as evaluate the heterogeneity in $PM_{2.5}$ mortality rate across different counties in the US. We also attempt to identify whether environmental quality index modifies the association between $PM_{2.5}$ and total non-accidental death risk.

Significant statement

A better understanding of effect modification on the relative mortality rate may help inform local air quality efforts to best alleviate the adverse health risk of air pollution. It can also have implications for policy-makers so that more stringent strategies can be taken to reduce air pollution and the corresponding mortality risk.

Chapter II

Literature Review

By 1990s, time-series studies conducted at a single location [12-15], demonstrated that even very low concentration of air pollution levels were associated with mortality rates in cities in the United States, Europe, and other developed regions. However, the inconsistency in findings of single-site studies on this association leads to the critique and questioning of the choice of particular cities. From then on, a hierarchical modeling approaches which can combine information across cities have been applied.[16]

While the epidemiological evidence relating the association between exposure to $PM_{2.5}$ and total mortality is substantial, there has been limited studies of effect modifications on this association. The challenge in understanding effect modification lies in the huge heterogeneity among study designs and populations, with a variety of health outcomes, pollutants, confounders, regions, and effect modifiers.[17] Levy JI et al.[10] reported that the regional differences in PM mortality risk estimates have often been attributed to factors such as geographic variability in particle composition and spatial heterogeneity of constituents. Baxter LK et al.[9] demonstrated that that the difference between cities may be due to differences in the air pollution mixture as well as differences in exposure (i.e., exposure measurement error or city-specific exposure factors), and possibly those related to local sources such as traffic.

Chapter III

Methods

Data

Core Based Statistical Area (CBSA) was defined as centers of an urban region of at least 10,000 people and adjacent areas that are socioeconomically tied to the urban center by commuting.[18] A CBSA may consists of one or more than one counties. Daily mortality rates as well as environmental quality indexes for each county were obtained from Environmental Protection Agency (EPA). The relative risks of daily $PM_{2.5}$ levels and total non-accidental death model with various lag structures were obtained from Poisson log-linear models estimated for each county separately. The mortality rate was measured as the log relative risk of total non-accidental death in counties per unit increase in $PM_{2.5}$. Since we would look at the association between $PM_{2.5}$ and total non-accidental death risk on both county-data and CBSA data, for the CBSA analysis, we removed the counties that are not part of a CBSA. The total number of counties is 433, while the number of counties belonging in a CBSA is 277.

Statistical Models

We examined the association between $PM_{2.5}$ and total non-accidental death risk as well as evaluate the heterogeneity in $PM_{2.5}$ mortality rate across different counties using a two-stage Bayesian hierarchical modeling approach. To identify whether environmental quality index modifies the association between $PM_{2.5}$ and total non-accidental death risk on county-data and CBSA data, explanatory variables were added in the second-stage.

First, we estimated the overall pooled effects of the mortality risks in a Bayesian regression model. Here β_c denotes the true log relative risk at a given county, μ denotes the pooled effect estimates of the log relative risk and σ^2 denotes the between-county variability.

$$\text{Model 1:} \quad \beta_c = \mu + \varepsilon_c \quad \varepsilon_c \sim N(0, \sigma^2)$$

Second, we fit a Bayesian hierarchical model to decompose the variability into between-CBSA variability and between-county variability. Here β_{cj} denotes the true log relative risk at county c within CBSA j, μ denotes the estimates of the mean log relative risk, σ^2 denotes the between-county variability after accounting for CBSA effects, and τ^2 denotes the between-CBSA variability.

$$\text{Model 2:} \quad \beta_{cj} = \mu_j + \varepsilon_{cj} \quad \mu_j \sim N(\mu, \tau^2) \quad \varepsilon_{cj} \sim N(0, \sigma^2)$$

Third, to look at the effect modification of environmental quality index, we added an explanatory variable to Model 1, where Z_c is the county-specific environmental quality index. It measures the change in the true log relative risk associated with1-unit increase in the corresponding county-specific environmental quality index.

$$\text{Model 3:} \quad \beta_c = \mu + \alpha Z_c + \varepsilon_c \qquad \varepsilon_c \sim N(0, \sigma^2)$$

Fourth, to look at the effect modification of environmental quality index account for CBSA heterogeneity, we added an explanatory to model 2, where Z_{cj} is the CBSA-specific environmental quality index. It measures the change in the true relative morality risk associated with1-unit increase in the corresponding county-specific environmental quality index after accounting for CBSA.

$$\text{Model 4:} \quad \beta_{cj} = \mu_j + \alpha Z_{cj} + \varepsilon_{cj} \qquad \mu_j \sim N(\mu, \tau^2) \qquad \varepsilon_{cj} \sim N(0, \sigma^2)$$

The specification of these Bayesian model is completed with the following prior distributions. We used the R package R2WinBUGS and WinBUGS to perform Markov chain Monte Carlo (MCMC). The total number of MCMC iterations is 40000 with a burn in of 20000 samples. The prior distributions are:$\mu \sim N(0,1000000)$, $\alpha \sim N(0,1000000)$, $\sigma^2 \sim unif(0, 100)$, $\tau^2 \sim unif(0, 100)$.

Chapter IV

Results

Table 1 summarizes the descriptive statistics of the environmental quality indexes. We can see that the average EQIs of county-data and CBSA data are nearly the same.

On average across counties, the pooled change in relative risk per 10 unit increase in $PM_{2.5}$ levels for lag time 0, 1 and 2 are: 0.1470(-0.0425, 0.3252), 0.3142(0.1159, 0.4638), 0.3889(0.2359, 0.5393). For the CBSA analysis, the overall estimates of log relative risk for lag 0, 1 and 2 are: 0.0392(-0.1973, 2.3690), 0.2071(0.0126, 0.0407), 0.3334(0.0865, 0.5736). Hence, the overall estimates obtained from considering all counties are slightly higher than that obtained from using only counties within a CBSA.

Figure 1 shows the estimated effect modification of log relative risks with 95% posterior intervals (PI) for different environmental quality indexes by exposure day lag under model 3 and model 4. As figure 1 indicates, the air and built environment domains are significant positive effect modifiers while land, water and sociodemographic domain are non-significant. Also, as the lag increases from 0 to 2, the air domain become less significant while build domain are more significant. The effect modification slope by the air domain during lag 0, 1, 2 are: 0.4398(95%PI: 0.1115,0.7704), 0.2676(95%PI:-0.0466, 0.5990), -0.0755(95%PI:-0.3624, 0.1729). This translates to: for current day log relative risk(lag=0), one-unit increase in the environmental quality index of the air domain will leads to a 0.44% increase in the county-specific log relative risk on mortality per 10-unit increase in $PM_{2.5}$. For the CBSA-only analysis, the effect of air domain for lag 0, 1, and 2 are: 0.5215(0.0173, 0.9679), 0.5021(0.0918, 0.8769). For the built environment domain during lag 0, 1, and 2, the effect modifications are: 0.4519(95%PI: 0.0310, 0.8662), 0.4977(95%PI: 0.2526, 0.7519), 0.5661(95%PI: 0.222, 0.8686) in the all-county analysis. For the CBSA-only analysis, the corresponding estimates are: 0.4396(95% PI:-0.0583, 0.9272), 0.6852(95% PI: 0.2546, 1.1180), 0.5679(95% PI: 0.2732, 0.9240).

We then examined the robustness of the air and built environment domain effect modification after adjustment by other environmental indexes. Figure 2 and Figure 3 show mean estimates and 95% posterior intervals of air and built environment domain after adjusting for the other four environmental quality indexes, one at a time. Overall, the effect modifications by air and built environment domain are similar to those without adjustment.

Chapter V

Discussion

According to the Figure 1, we found that the CBSA estimates and the county estimates are very similar. However, the CBSA estimates have wider confidence interval than the estimates of county-data. The reason that CBSA estimates have wider confidence interval may due to two reasons: First, we removed counties that are not within a CBSA, leading to a smaller sample size, thus larger confidence interval. Second, this may due to a large heterogeneity variance.

Among all the five environmental quality index domain, air and built environment are significant associated with higher $PM_{2.5}$ risks. Considering that highway safety, highway proportion and main roads proportion are included in the built environment domain, it may be reflecting $PM_{2.5}$ compositions that are associated with higher levels of traffic-related pollutants, as suggested by Baxter et al[9]. As to the air domain, since the index includes the level of nitrogen dioxide, PM_{10} and $PM_{2.5}$ itself, it suggests that long-term poor air quality is associated with higher $PM_{2.5}$ risks on the mortality.

Reference

[1]Martinelli N, Olivieri O, Girelli D. Air particulate matter and cardiovascular disease: a narrative review. Eur J Intern Med. 2013 Jun;24(4):295-302.

[2]Esworthy R. Air quality: EPA's 2013 changes to the particulate matter (PM) standard. Congressional Research Service 7-5700, n. R42934; 2013. p. 6.

[3]Juda-Rezler K, Reizer M, Oudinet JP. Determination and analysis of PM_{10} source appor- tionment during episodes of air pollution in Central Eastern European urban areas. The case of wintertime 2011, 45. ; 2011. p. 6557–66.

[4]World Health Organization (WHO). Health effects of particulate matter. Policy implications for countries in eastern Europe, Caucasus and central Asia. Copenhagen: WHO Regional Office for Europe; 2013 [http://www.euro.who.int/__data/assets/pdf_file/0006/189051/ Health-effects-of-particulate-matter-final-Eng.pdf, accessed 27 August 2013].

[5]Kim KH, Kabir E, Kabir S. A review on the human health impact of airborne particulate matter. Environ Int. 2015 Jan;74:136-43.

[6]Pelucchi C, Negri E, Gallus S, Boffetta P, Tramacere I, La Vecchia C. Long-term particulate matter exposure and mortality: a review of European epidemiological studies. BMC Public Health. 2009 Dec 8;9:453.

[7]Shah ASV, Langrish JP, Nair H, McAllister DA, Hunter AL, Donaldson K, et al. Global asso- ciation of air pollution and heart failure: a systematic review and meta-analysis. The Lancet 2013;382:1039–48.

[8]Chuang K-J, Yan Y-H, Chiu S-Y, Cheng TJ. Long-term air pollution exposure and risk factors for cardiovascular diseases among the elderly in Taiwan. Occup Environ Med 2011; 68:64–8.

[9]Baxter LK, Duvall RM, Sacks J. Examining the effects of air pollution composition on within region differences in $PM_{2.5}$ mortality risk estimates. J Expo Sci Environ Epidemiol. 2013 Sep-Oct;23(5):457-65.

[10]Levy JI, Diez D, Dou Y, Barr CD, Dominici F. A meta-analysis and multisite time-series analysis of the differential toxicity of major fine particulate matter constituents. Am J Epidemiol 2012; 175: 1091–1099.

[11]Messer LC, Jagai JS, Rappazzo KM, Lobdell DT. Construction of an environmental quality index for public health research. Environ Health. 2014 May 22;13(1):39.

[12]Health effects of outdoor air pollution. Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. Am J Respir Crit Care Med 1996; 153:3–50.

[13]Lipfert F, Wyzga R. Air pollution and mortality: issues and uncertainty. J Air Waste Manag Assoc 1993;45:949–66.

[14]Li Y, Roth H. Daily mortality analysis by using different regression models in Philadelphia county, 1973–1990. Inhal Toxicol 1995;7:45–58.

[15]Schwartz J. Air pollution and daily mortality in Birmingham, Alabama. Am J Epidemiol 1995;137:1136–47.

[16]Dominici F. Invited commentary: air pollution and health-what can we learn from a hierarchical approach? Am J Epidemiol. 2002 Jan 1;155(1):11-5; discussion 16.

[17]Bell ML, Zanobetti A, Dominici F. Evidence on vulnerability and susceptibility to health risks associated with short-term exposure to particulate matter: a systematic review and meta-analysis. Am J Epidemiol. 2013 Sep 15;178(6):865-76.

[18]Source available at http://en.wikipedia.org/wiki/Core_Based_Statistical_Area

Appendix

Figure1. Posterior mean and 95% posterior intervals for the effect modification of of different environmental quality index on $PM_{2.5}$ relative risks by lag time. The relative risks are per 10 unit increase in $PM_{2.5}$ on non-accidental mortality. The analysis are conducted using either all counties or counties within a CBSA.

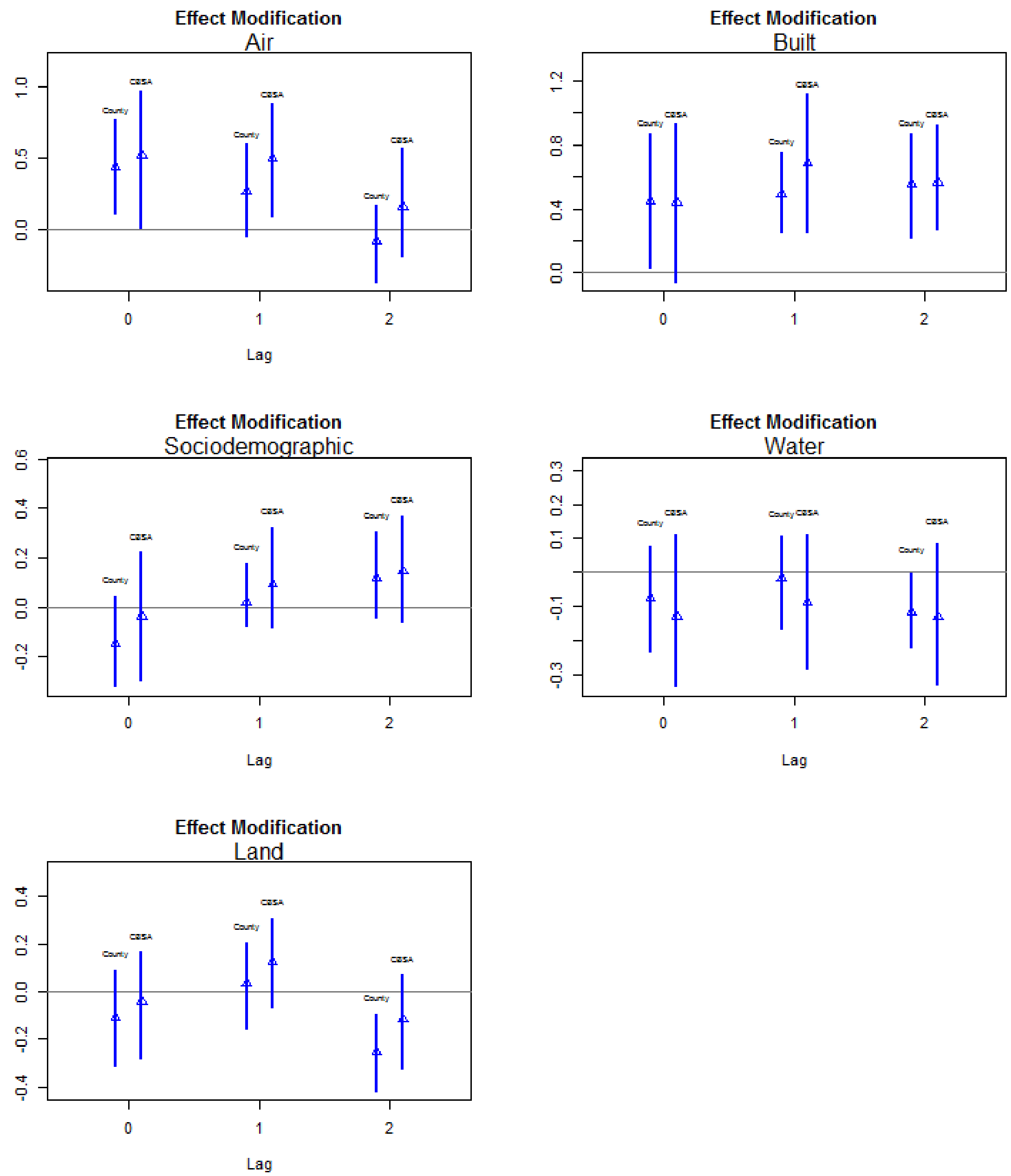


Figure2 . Posterior mean and 95% posterior intervals of the effect modification by the air index after adjusting for other environmental quality index

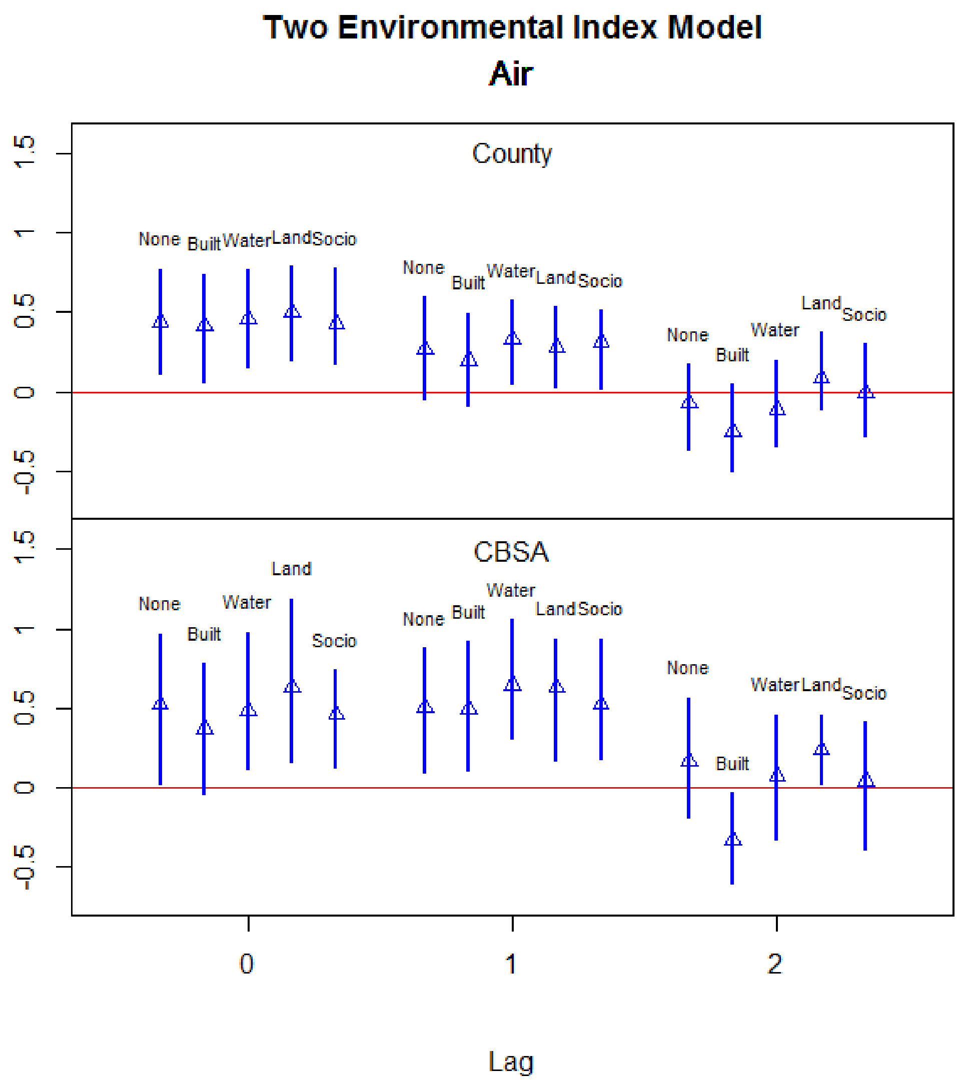
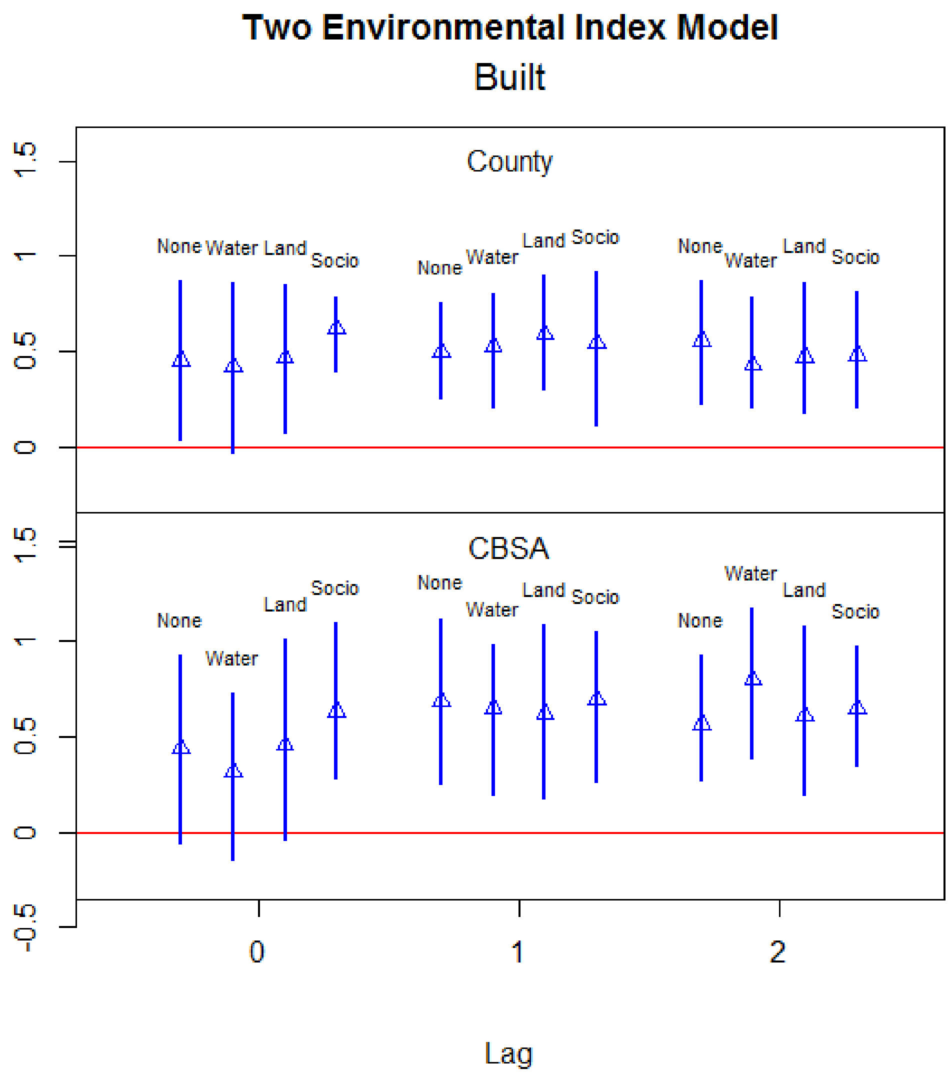


Figure 3. Posterior mean and 95% posterior intervals of the effect modification by the built environment index after adjusting for other environmental quality index



	Median(min, max), SD	
EQIs	All	CBSA
Air	1.1637(-1.506, 2.7898), 0.5887	1.0662(-1.506, 2.7898), 0.5995
Built	0.7918(-2.0637, 2.615), 0.4535	0.7478(-2.0637, 2.615), 0.4589
Land	0.5084(-5.1192, 2.0945), 0.8456	0.522(-5.1192, 1.7299), 0.8859
Water	0.7844(-1.5971, 1.4782), 1.045	0.8417(-1.5676, 1.4248), 0.9973
Sociodemographic	0.5092(-3.1729, 3.5311), 0.9234	0.4662(-3.1729, 3.5311), 0.9484

Table1. The summary statistics(median, minimum, maximum and standard deviation) of environmental indexes

WinBUGS code:

Model1:

```
model{
   for (i in 1:N) {
      y.hat[i] ~ dnorm(y[i], prec2[i])
      y[i] ~ dnorm(mu[i], tau2)
           mu[i] <- alpha
   }
   #Priors
   alpha ~ dnorm(0.00000E+00, 1.00000E-06)
   sigma2 ~ dunif(0.00000E+00, 100)
   tau2 <- pow(sigma2, -2)
}
```

Model2:

```
model{
for( i in 1:N){
      y.hat[i] ~ dnorm(y[i], prec2[i])
```

```
      y[i] ~ dnorm(mu[i], tau2)
           mu[i] <- mur[ID[i]]
   }
for(i in 1:n){
     mur[i] ~ dnorm(alpha, gamma2)
}
   #Priors
   alpha~dnorm(0.00000E+00, 1.00000E-06)
   sigma2 ~ dunif(0.00000E+00, 100)
   tau2 <- pow(sigma2, -2)
   eta2 ~ dunif(0.00000E+00, 100)
   gamma2 <- pow(eta2, -2)
}
```

Model3_air:

```
model{
   for (i in 1:N) {
      y.hat[i] ~ dnorm(y[i], prec2[i])
      y[i] ~ dnorm(mu[i], tau2)
           mu[i] <- alpha+air_EQI_22July2013[i]*air
   }
   #Priors
   alpha ~ dnorm(0.00000E+00, 1.00000E-06)
   sigma2 ~ dunif(0.00000E+00, 100)
   tau2 <- pow(sigma2, -2)
   air ~ dnorm(0.00000E+00, 1.00000E-06)
}
```

Model4_air_CBSA:

```
model{
   for (i in 1:N) {
      y.hat[i] ~ dnorm(y[i], prec2[i])
      y[i] ~ dnorm(mu[i], tau2)
    mu[i]<- mur[i] + air_EQI_22July2013[i]*air
           mur[i] <- murs[ID[i]]
   }
for(i in 1:n){
     murs[i] ~ dnorm(alpha, gamma2)
}
   #Priors
   alpha ~ dnorm(0.00000E+00, 1.00000E-06)
   sigma2 ~ dunif(0.00000E+00, 100)
   tau2 <- pow(sigma2, -2)
   eta2 ~ dunif(0.00000E+00, 100)
   gamma2 <- pow(eta2, -2)
   air ~ dnorm(0.00000E+00, 1.00000E-06)
```

```
  icc<-gamma2/(gamma2+tau2)
}
```