2013-01-01

Functional Data Analysis to Guide a Conditional Likelihood Regression in a Case-Crossover Study Investigating Whether Social Characteristics Modify the Health Effects of Air Pollution

Juana Maribel Herrera Hernandez
University of Texas at El Paso, maribelhh@gmail.com

FUNCTIONAL DATA ANALYSIS TO GUIDE A CONDITIONAL LIKELIHOOD REGRESSION IN A CASE-CROSSOVER STUDY INVESTIGATING WHETHER SOCIAL CHARACTERISTICS MODIFY THE HEALTH EFFECTS OF AIR POLLUTION

JUANA MARIBEL HERRERA HERNANDEZ

Department of Mathematical Sciences

APPROVED:

Joan G. Staniswalis, Chair, Ph.D.

Sara E. Grineski, Ph.D.

Amy E. Wagler, Ph.D.

Benjamin C. Flores, Ph.D.
Dean of the Graduate School

to my

SON, HUSBAND and MOTHER

with love

FUNCTIONAL DATA ANALYSIS TO GUIDE A CONDITIONAL LIKELIHOOD REGRESSION IN A CASE-CROSSOVER STUDY INVESTIGATING WHETHER SOCIAL CHARACTERISTICS MODIFY THE HEALTH EFFECTS OF AIR POLLUTION

by

JUANA MARIBEL HERRERA HERNANDEZ

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

August 2013

Acknowledgements

I was privileged to work with Dr. Joan Staniswalis who patiently took me through the difficult task of research. Dr. Staniswalis, your support, encouragement and guidance were essential to my success here, thank you for sharing your knowledge and experience with me, it is something I'll remember forever.

I would like to acknowledge Dr. Sara Grineski for letting me be part of this project and serving on my comittee, all our meetings during the past year were very helpful in achieving this goal. Dr. Wagler, you also served as my committee member, I really appreciate that you read my work and made valuable comments that improved it.

I would like to thank Priyangi Bulathsinhala for being so generous as to provide me with a copy of her work. My sincere appreciation goes to the members of the Department of Mathematical Sciences: the faculty, staff and students, you really made my stay in UTEP one I'll always remember.

My work was supported by the National Institutes of Health (NIH) score grant ISC3GM094073-01.

Finally, I would like to thank my mother and my husband; without their love and support, I am not sure whether I would have even been able to accomplish this goal. You encouraged, helped and cheered whenever I needed.

Abstract

In this study we are focused on exploring whether social characteristics modify the relationship between air pollution and hospitalizations due to asthma or chronic pulmonary obstructive desease (COPD) in El Paso, Tx. The case-crossover design with conditional regression analysis was used, here the controls and the case are the same subject at different times and has the advantage of removing confounding by permanently fixed factors. Social charateristics are included in the models as interactions with the pollutants, variables included are age, sex, ethnicity and insurance status as indicator for the socio-economic status. The pollutant's lags were chosen using the historical functional linear model to estimate the association between the response and pollutant at all lags simultaneously. The regression coefficient function was calculated by P-splines with the smoothing parameter chosen with a modified ridge trace method. We included single pollutant analyses for NO_2 and $PM_{2.5}$ for both asthma COPD diseases, adjusting for apparent temperature (combination of temperature and dew point) and wind speed. The lags for low and high wind speed were chosen, in the case of asthma, based on previous literature and in the case of COPD based on odds ratios. Subgroup analyses by ethnicity are presented, in order to compare Hispanics and Non-Hispanics without the assumption that the weather variables have the same effect for all subgroups. We found that when $PM_{2.5}$ is equal to the 98% percentile of the daily values, Hispanics are more likely to be hospitalized due to asthma or COPD than Non-Hispanics, but when NO_2 is equal to the 98% percentile for the daily values, contrary to $PM_{2.5}$, Non-Hispanics are more likely to be hospitalized due to asthma or COPD than Hispanics. For children with Medicare the probability of being hospitalized for asthma increases significantly when NO_2 increases by one interquartile range. This exploratory study is looking for patterns that can be later compared with findings in other cities as part of a comprehensive review.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

An association between air pollutants and health outcomes has been found in many studies. Zanobetti et al.(2006), reported, a significant association between NO_2, $PM_{2.5}$, and Black Carbon and the risk of hospitalization for myocardial infarction. They also, reported a significant association between Black Carbon, $PM_{2.5}$ and Carbon Monoxide (CO) and the risk of hospitalization for pneumonia. That study was conducted in Boston, MA for all citizens and permanent residents aged 65 and over with Medicare coverage.

A study by Ostro et al.(2006) was conducted in 9 California counties; the results of a single pollution model indicated evidence of significant association of particles of 2.5 micrometers in diameter or smaller ($PM_{2.5}$) with cardiovascular and respiratory causes of mortality for all ages. Studies outside of the USA also found a significant association between air pollutants and health effects. Tenías et al. (1998), concluded that the concentrations of ambient of NO_2 and O_3 in Valencia, Spain were significantly associated with emergency room visits for asthma among people over 14 years of age. More studies can be found in the following United States Enviromental Protection Agency (EPA) documents: *Provisional Assessment of Recent Studies on Health Effects of Particulate Matter Exposure (2006)* [32] and *Integrated Science Assessment for Oxides of Nitrogen Health Criteria (Final Report)* [31].

In Staniswalis et al.(2005), a significant association between evening PM10 exposures and nonaccidental daily mortality is found in El Paso from 1992 to 1995. Other studies adressing health effects of air pollution in El Paso, TX are Staniswalis et al. (2009) and Grineski et al.(2011). Staniswalis et al. (2009) found a significant association between an increase of $10\mu g/m^3$ in the hourly $PM_{2.5}$ across a day and the daily mortality (considering

only natural deaths) with relative risk of 1.017(95% CI: 1.001-1.034), 1.024(95% CI: 1.012-1.037), 1.016(95% CI: 1.006-1.027) for the same day and previous two days respectively. Grineski et al. (2011) concluded that there were respiratory health effects associated with dust and low wind events in El Paso. The relative risks of hospitalization for asthma three days after a low wind event and of hospitalization for acute brochitis one day after a dust event in children (aged 1-17) were 1.19(95% CI: 1.00-1.41) and 1.33(95% CI: 1.01-1.075) respectively.

Whether social characteristics modify the relationship between air pollution and health is another question that researchers have examined [5] [9] [16]. In Grineski et al.(2010), social characteristics such as race, ethnicity and insurance status were examined as effect modifiers between NO_2 and children's asthma hospitalizations in Phoenix, Arizona(US) between 2001 and 2003. Significant findings were found: the risk of hospitalization for asthma in children without insurance is 1.39(95% CI: 1.1-1.8) higher than those with private insurance and 1.38(95% CI: 1.06-1.80) higher than those with Medicaid at exceedances of 0.02 parts per million (ppm) that is the 95th percentile of NO_2 desviations from the seasonal mean. Among children without insurance, Hispanic children have higher risk of hospitalization than white children. The risk of hospitalization for asthma in Hispanic children without health insurance is higher than Hispanic children with private insurance or Medicaid.

This thesis continues to examine ethnicity and race as possible modifiers of the effects of air pollution on respiratory hospitalizations in El Paso. The objectives of this study are as follows:

1. Determine if socio-economic status (SES) modify the effects of daily air pollution on respiratory hospitalizations.

2. Determine if Hispanics, as compared to non-Hispanics, face increased odds of respiratory hospitalizations due to increases in air pollution in El Paso.

3. Determine SES modify the effect of pollution differently for Hispanic as compared to

non-Hispanic in El Paso.

Insurance status is used as a sugorrate for SES since that is all that is available in the data.

The following air pollutants are included in this work:

Particulate matter

Particulate matter (PM) or particle pollution "is a complex mixture of extremely small particles and liquid droplets. Particle pollution is made up of a number of components, including acids (such as nitrates and sulfates), organic chemicals, metals, and soil or dust particles", according to EPA website [33]. Studies on health effects of particulate matter exposure are concentrated on particles of 10 micrometers in diameter or smaller because those can be inhaled and enter the lungs and bloodstream [36]. The notation for particles of $10\mu m$ or smaller is PM_{10} and for particles of $2.5\mu m$ or smaller is $PM_{2.5}$

Nitrogen dioxide (NO_2)

NO_2 is a highly reactive gas. It forms from most combustion processes using an air as oxidant, for example, internal combustion engines and power plants. Nitrogen dioxide contributes to the formation of ground level ozone and fine PMs, which are 2.5 micrometers in diameter and smaller. [34].

Ozone O_3

In the upper regions of the atmosphere, the ozone layer protects life from the sun's UV rays. Ground level ozone is an atmospheric pollutant. Ground level ozone is created by chemical reactions between oxides of nitrogen (NOx) and volatile organic compounds (VOC) and can be transported long distances by wind. [35]

1.1 Design and Analysis

The case-crossover design with analysis by conditional logistic regression is used to study the association between air pollution and health outcomes. Developed by Maclure (1991), the case-crossover design is a special case of the case-control design in which the controls and the case are the same subject at differents times. Self-matching of cases simplifies control selection, and removes confounding by permanently fixed factors (e.g. ethnicity, race, gender).

Matching in a case-crossover study can be summarized as follows [17]. The population is divided into strata, and there is only one individual per stratum (e.g. the number of individuals is equal to the number of strata). If a stratum has no case then is not sampled (e.g. if the individual is not a case, then the strata is ignored). If the stratum has a case then t_i is assigned as the time event and a set of M referent times are determined and we assume that the individual is not a case at the referent times. For example, control times can be chosen to match on the day of the week, in the same month and year as the event day. The case and its controls form a 1:M matched case-control set. The matched sets from different individuals are then analysed by conditional logistic regression, as in a standard matched case-control study.

In the case-crossover design the problem is to estimate the slope parameters (log-odds) for each covariate, however an intercept parameter is generated for each stratum. If we have n strata and p covariates then we need to estimate $n + p$ parameters, but we cannot estimate the parameters by maximum likelihood estimator (mle) because the number of parameters increases as the sample increases. With conditional logistic regression, the nuisance parameters can be conditioned out of the analysis, and a conditional likelihood function is created. This is explained in detail in Chapter 2.

In this thesis, air pollutants and weather conditions are used as covariates; we cannot assume that the health outcomes depend only on the conditions at the time of hospitalization. Lagged values of the pollution and weather conditions need to considered as well,

although choosing the best lag time to predict health outcomes is problematic. In Belleudi et al.(2010), eleven conditional logistic regression analysis with a case-crossover design were performed to study the association between PM2.5 and hospitalizations for acute coronary syndrome. They explored the individual lag effects of PM2.5 from day 0 to day 6, and then by taking the average over lag 0 through 1 day, lag 0 through 2 days, and lag 0 through 5 days. Reporting on eleven differents models is time consuming and to reach a conclusion we have to deal with multiple comparisons. In Staniswalis et al.(2005), weather variables with 3 lag times were included in the model and then a backward variable selection procedure was performed to keep variables in the model that were significant at the 0.2 level of significance. Grineski et al.(2011) used the highest odds ratio to select lag times for dust and low wind events. We deal with the problem of lag-time specification for NO_2 and $PM_{2.5}$, by fitting the historical functional linear model to incorporate the hourly pollutant measures into the regression model with a continuous lag, instead of fixed time-lag of daily averages [4]. The procedure described in Chapter 3 is:

- Preprocess the data to obtain the B-splines functions as described in Chapter 3 section 4.
- We used a modified version of the COXPH function in the R-package to get the coefficients for the continuous lag.
- The ridge trace is used to guide the choice of the smoothing parameter in the non-parametric functional linear model.
- The slope function is estimated for the pollutants.
- The lag-time is chosen based on the shape of the slope function.

The data, statistical methods and results are reported in Chapter 4. Finally, the conclusions are presented in Chapter 5.

Chapter 2

Conditional Logistic Regression

2.1 Logistic Regression

When the response Y is binary (e.g. whether an individual was hospitalized or not), we can define Y as

$$Y_i = \begin{cases} 1 & \text{if the } i_{th} \text{ individual is a case} \\ 0 & \text{if the } i_{th} \text{ individual is not a case.} \end{cases}$$

Now Y_i is a Bernoulli random parameter with $E[Y_i] = \pi$. Since $0 \leq E[Y_i] \leq 1$, the linear regression model is not appropiate in this case because it does not satisfy that constraint. The logistic regression model is used when the response is binary. The model is [10]:

$$E[Y|x] = \pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \ , \tag{2.1}$$

where Y is the response and x is an independent variable. The logit transformation of $\pi(x)$ is:

$$g(\pi) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x$$

The logit of $\pi(x)$ is linear in the parameters (α, β), continuous and its range is $-\infty$ to ∞. The method of maximum likelihood is used for estimation of the parameters α and β. We need the likelihood function. Suppose we have n independent observations of the pair (y_i, x_i), $i = 0, ..., n$, where y_i is the values of the response variable for the i_{th} subject, and x_i is the values of the independent variable for the i_{th} subject. The distribution of $y_i|x_i$ is independent Bernoulli$(\pi(x_i))$, hence the likelihood function is:

$$l(\alpha, \beta; Y|X) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i}$$

Using software we can obtain the estimates for the parameters given data Y and X [10]. However, this is possible only when the number of parameters is constant and the sample is large. When the number of parameters is bigger than the sample size we can use the conditional logistic regression to avoid the estimation of any nuisance parameter. Hence, we need to buid a conditional likelihood function.

2.2 Conditional Likelihood

The maximum likelihood method is widely used for estimation due to its consistency and asymptotic properties. When we have the presence of nuisance parameters, a conditional likelihood function can be used instead of the complete likelihood function. In Kalbfleisch and Sprott (1970) we can find two conditions for existence of the conditional likelihood estimator and no loss for estimation of information of the parameter of interest. The two conditions are given below:

1. The distribution function of the observations factors into two distribution functions.

$$f(\underline{x}; \beta, \alpha) \prod dx_i = \left(f(\underline{x}; \beta|\underline{T}) \prod dx_i / \prod dT_i \right) \times \left(f(\underline{T}; \beta, \alpha) \prod dT_i \right) \qquad (2.2)$$

   where $\underline{x} = (x_1, ..., x_n)$ are our observations, β is the parameter of interest and α is the nuisance parameter. $\underline{T} = (T_1, ..., Tk)$ are jointly sufficient for α when β is known.

2. The second factor on the right side of (2.2) contains no available information concerning β when α is unknown. Then we can use the first factor on the right side of (2.2) with no concerns about loss of information.

If the two conditions are satisfied the conditional likelihood of β is proportional to $f(\underline{x}; \beta|\underline{T})$ $\prod dx_i / \prod dT_i$. Here $\prod dx_i / \prod dT_i$ represents the infinitesimal volume within the subspace

of $\mathbb{R}^n$ obtained by setting T_i = constant, $(i = 1, ..., k)$ [12]. Sprott and Kalbfleisch (1969) derived an expression for $\prod dx_i / \prod dT_i$ using methods for handling tranformations of variables. If $\underline{T}$ is not function of β, then we can ignore the quantity $\prod dx_i / \prod dT_i$.

In the presence of nuisance parameters, sometimes it is convenient to calculate the cmle instead of the maximum likelihood estimator (mle). The question is whether the properties of the cmle are the same as the properties of the mle, or how to measure or assess the differences between both methods. Discussions about the properties of the cmle can be found in Andersen (1970), Huque & Katti (1976), Basawa (1981), Liang (1983,1984) and Sartori & Severini (2004). Andersen(1970) proved that under a certain set of conditions the cmle $\hat{\beta}_c$ converges almost surely to the true value β_0 and its asymptotic distribution is normal with mean β_0 and a specified variance. If we are conditioning on T, where T is a sufficient statistic for α and an ancillary statistic for β, then the variance of $\hat{\beta}_c$ asymtotically attains the Cramer-Rao lower bound. For exponential families it is not necessary that T be an ancillary statistic, the lower bound is asymptotically attained if and only if T is weakly ancillary with respect to β [1]. Andersen(1970) introduced the term weakly ancillary:

Definition 1. *Let $(\beta, \alpha) \in \Omega_1 \times \Omega_2$ the parameter space. Let $G_{\beta,\alpha}(t)$ be the marginal distribution of T, the sufficient statistic for α. If for any given set of values $(\beta_0, \alpha_0) \in \Omega_1 \times \Omega_2$ and for any other value $\beta \in \Omega_1$ there exists a point $\alpha = \alpha(\beta) \in \Omega_2$, such that*

$$G_{\beta,\alpha}(t) \equiv G_{\beta_0,\alpha_0}(t)$$

then T is weakly ancillary for β.

It is not easy to establish when a statistic has this property of weak ancillarity. However, necessary and sufficient conditions when the distribution of X belongs to the exponential family have been provided by Andersen(1970) and Liang (1983). When t is not weakly ancillary for β, Liang (1984) introduces the concept of asymptotically weakly ancillary statistic:

Definition 2. *Consider the distributions $p(X;\tau)$, $g(T;\tau)$ and $f(X|T;\theta)$. A statistic T is called asymptotically weakly ancillary for β if, for each $\tau = (\beta, \alpha)$, $I_g(\tau)/I_p(\tau) \to 0$ as $n \to$*

∞, *or equivalently* $I_f(\tau)/I_p(\tau) \to 1$ *as* $n \to \infty$. *Here* $I_p(\tau) = \inf_{c\in\mathbb{R}} E(\partial logp/\partial\beta - c\partial logp/\partial\alpha)^2$, $I_g(\tau) = \inf_{c\in\mathbb{R}} E(\partial logg/\partial\beta - c\partial logg/\partial\alpha)^2$ *and* $I_f(\tau) = \inf_{c\in\mathbb{R}} E(\partial logf/\partial\beta - c\partial logf/\partial\alpha)^2$.

This means that the information contained in T is small compared to the full information when the sample size becomes large. Liang (1984) proved if a certain set of conditions are met and T is asymptotically weakly ancillary statistic then the cmle estimator is asymptotically equivalent to the unconditional maximum likelihood estimator. Here is an example from Liang (1984) p.311; Consider $Y = (Y_1, .., Y_n)$, where $Y_i \sim Bernoulli(1, \pi_i)$ such that $\text{logit}(\pi_i) = \alpha + \beta X_i$ with the X_i's being fixed. The full likelihood is

$$p(Y;\tau) = \exp(\beta \sum Y_i X_i + \alpha \sum Y_i) \prod \{1 + \exp(\alpha + \beta X_i)\}^{-1},$$

with $T = \sum Y_i$. The conditional likelihood is

$$f(\sum Y_i X_i = w | T = t) = \frac{N_{w,t} e^{\beta w}}{\sum_u N_{u,t} e^{\beta u}},$$

where $N_{w,t}$ is the number of distinct ordered sets of t numbers, taken from $X_1, ..., X_n$ whose sum is w, and $\sum_u$ is the summation over all values taken by the random variable $\sum Y_i X_i$. It is assumed that for large samples the X_i's are generated from some unknown distribution with finite moments of every order. In this case Liang (1984) concluded that the conditions are met and so the cmle estimator is asymptotically equivalent to the unconditional maximum likelihood estimator.

2.3 Conditional Likelihood Applied to the Case-Crossover Study

Let $Y = (y_1, ..., y_n)$, where $y_i|x_i$ are independent Bernoulli(π_i) such that $\log(\pi_i) = \alpha + \beta x_i$ where the x_i's are the risk factors. In a prospective study we are interested in the probability of $\pi(x_i)$ as defined in equation (2.1). In a case control design the probability of interest is $\pi^*(x_i) = P(x_i|y_i = 1, s_i = 1)$, where s_i is defined by the following:

$$s_i = \begin{cases} 1 & \text{if the } i_{th} \text{ subject is selected in the sample} \\ 0 & \text{if the } i_{th} \text{ subject is not selected in the sample.} \end{cases}$$

Then $\pi^*(x_i)$ is the probability that the sampled case was exposed to the risk factors. $\pi^*(x_i)$ is proportional to the logistic regression model $\pi(x_i)$ with the same effect parameter β as with the prospective study, but with a new intercept α^* [21]. The full conditional likelihood function is:

$$l(\alpha^*, \beta) = \prod_{i=1}^{n} \pi^*(x_i)^{y_i}(1 - \pi^*(x_i))^{1-y_i} = e^{\alpha^* \sum y_i + \beta \sum y_i x_i} \prod (1 + e^{\alpha^* + \beta x_i})^{-1} \tag{2.3}$$

where $\alpha^* = \ln(\rho_1/\rho_0) + \alpha$. Assuming $\rho_0 = P(s_i = 1|y_i = 0, x_i) = P(s_j = 1|y_j = 0)$, and $\rho_1 = P(s_i = 1|y_i = 1, x_i) = P(s_j = 1|y_j = 1)$, β is known and x_i is fixed, the factors in equation (2.3) that contain α^* depend on the sample Y only through the function $T = \sum y_i$. Thus by the Factorization Theorem T is a sufficient statistic for α^* given β and $x_1, x_2, ..., x_n$.

In Chapter 1 we explained the problem with the use of the mle method for estimation of the parameters in a case-crossover design. Basically we have more parameters than strata. Conditioning the full likelihood function on $T = \sum y_i$, the sufficient statistic for α^*, we avoid estimation of the α_j^*'s, the intercept parameters for each strata. In a case-crossover study, we have for each strata $m + 1$ observations, one case and m controls. Then the conditional likelihood equation of the i_{th} strata is:

$$l(\beta_i) = f(y_{i,1}, y_{i,2}, ..., y_{i,m+1}|T = \sum_{j=1}^{m+1} y_{ij} = 1) \tag{2.4}$$

Recall that the case is its own controls at differents times. Let t_i be the index time (e.g. the time when the subject is a case) for the subject i. A rearrangement can be done in order that t_i=1. The probability of t_i=1 is:

$$P_1 : (x_{i1}, Y = 1) \text{ and } (x_{ij}, Y = 0),\ j \in W_{t_i},\ j \neq 1$$

where W_{t_i} is a referent set for the i_{th} subject and includes the referent times and the index time, x_{ij} is the risk factor status for the subject i at time j. Then the conditional likelihood equation used for this case is [21],

$$l(\beta_i) = P(P_1|data) = \frac{e^{\beta_i x_{i1}}}{\sum_{j \in W_{t_i}} e^{\beta_i x_{ij}}}$$

and the conditional likelihood of β is $l(\beta) = \prod_{i=1}^{n} l(\beta_i)$, where n is the number of strata. The procedure PROC PHREG in SAS maximizes this likelihood over β.

According to the SAS manual [24], you can perform conditional logistic regression with the PHREG procedure by using the discrete logistic model and forming a stratum for each matched set. The TIES=DISCRETE option requests the discrete logistic model. For matched case-control studies with one case per matched set (1:m matching), you can use the default TIES=BRESLOW. Another option to specify is the strata variable. In the case-crossover analyses, each event is treated as a separate stratum, however, where exposure is shared, we can condition the analysis on calendar day instead of on each event. For example, when control times are matched on the day of the week, in the same month and year as the event day, if more than one event ocurrs on the same day then the cases and controls share the exposures. With STRATA=CalendarDay we have fewer strata and the computation time can be reduced [37]. Wang et al. (2011) performed simulations combining the two approaches for handing the ties (discrete and Breslow) with calendar day as the strata option. First, they simulated the number of cause-specific hospital admissions on day i (Y_i) over a 10-year period as a Poisson random variable:

$$Y_i \ \sim \ \text{Poisson}\left(\ln(\lambda_i) = \beta_0 + \beta_1 PM_i\right) \ ,$$

where PM_i represents $\text{PM}_{2.5}$ levels on day i, β_1 represents the hypothesized log odds ratio associated with a unit ($\mu g/m^3$) increase in $\text{PM}_{2.5}$ on the same day; β_0 was set to 5. The controls in the referent set were all days in the same year, month, and day of the week as the simulated case day, excluding the case day. The analysis was conditioned on calendar day and the estimation of association between the level of exposure on the same

day and the hopitalization rate was the objective. Their results are presented in Figure 2.1. As an applied example, they evaluated the association between $PM_{2.5}$ levels and the risk of hospitalization for congestive heart failure among Medicare beneficiaries residing in Chicago using the timestratified case-crossover design. The results are presented in Figure 2.2. Finally, they concluded that fitting a stratified Cox model with the "Breslow "option for handling tied failure times provides unbiased health-effects estimates in case-crossover studies with shared exposures, whereas fitting with the "discrete "option leads to estimates which can be biased away from the null hypothesis of no association by $22\% - 39\%$, even for small simulated relative risks.

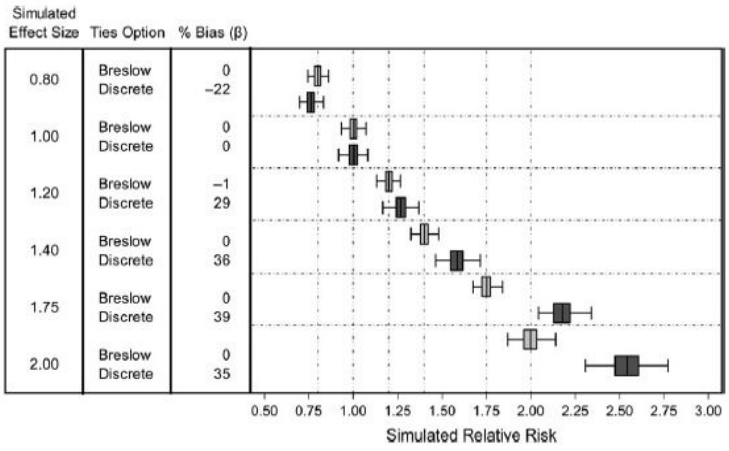


Figure 2.1: Results of analyses performed in SAS for simulated relative risks of hospitalization for heart failure ranging from 0.8 to 2.0 per 10 $\mu g/m^3$ increase in $PM_{2.5}$. Box plots show the distribution of effect estimates from 250 simulated data sets. Each box denotes the median value and interquartile range (25th to 75th percentiles), and whiskers denote the upper and lower adjacent values. The vertical dashed lines denote the simulated relative risk. *Figure taken from Wang et al. (2011).*

Table 2. Association Between Ambient $PM_{2.5}$ and Risk of Hospitalization for Congestive Heart Failure, Estimated With Different SAS[a] Procedures, Ties-Handling Options, and Frequency Weights, Among Medicare Beneficiaries Residing in Chicago, Illinois, 2000–2006

SAS Procedure	Ties Option	Conditioning On:	Weighted By:	Estimate, %[b]	95% Confidence Interval	Computational Time, seconds
PHREG	Breslow[c]	Each event		1.2	0.1, 2.4	18.3
PHREG	Breslow	Each day	Events/day	1.2	0.1, 2.4	0.8
PHREG	Discrete	Each day	Events/day	1.6	0.3, 2.9	9.3
LOGISTIC		Each event		1.2	0.1, 2.4	6.2
LOGISTIC		Each day	Events/day	1.6	0.3, 2.9	25.5

Abbreviation: $PM_{2.5}$, particulate matter with an aerodynamic diameter $\leq$2.5 μm.
[a] SAS Institute Inc., Cary, North Carolina.
[b] Percentage increase in risk of hospitalization for heart failure per 10-$\mu g/m^3$ increase in $PM_{2.5}$.
[c] Any of the available ties-handling options would yield identical results in this case.

Figure 2.2: Figure taken from Wang et al. (2011).

Chapter 3

Functional Linear Regression for Choosing the Lag

3.1 Ridge Regression

The simple linear regression model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \ , \tag{3.1}$$

where $\mathbf{Y}$ is the vector of the observed values, $\mathbf{X}$ is the predictors matrix, $\boldsymbol{\beta}$ is the coefficients vector and $\boldsymbol{\epsilon}$ is the error vector. The regression model (3.1) assumes that $E(\boldsymbol{\epsilon}) = 0$, $\sigma^2(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$, and that the ϵ_i are independent normal variables [13]. The normal equations derived by the method of least squares are

$$\mathbf{X}'\mathbf{X}\ \mathbf{b} = \mathbf{X}'\mathbf{Y}$$

therefore

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{3.2}$$

and $E[\mathbf{b}] = \beta$ and $\sigma^2(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Then

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} \ . \tag{3.3}$$

When the predictor variables are highly correlated, the estimated regression coefficients tend to have large sampling variability. Another effect of multicollinearity is that the interpretation of the regression coefficients as measuring the change in the expected value of the response variable when the given predictor variable is increased by one unit while

all other predictor variables are held constant is not fully applicable [13]. For example, if age and height of children are predictors variables, it is not realistic to vary age and hold height constant.

Ridge regression is a method to remedy multicollinearity problems. In this model the estimation of β by the method of least squares is modified to allow biased estimators of regression coefficients, these estimators are more accurate than the ones obtained by equation (3.2) (e.g. estimators from ridge regression have a larger probability of being close to the true parameter). See Figure 3.1. In ridge regression, we minimize a penalized sum of squares:

$$\sum_{i=1}^{n}(\mathbf{Y_i} - \hat{\mathbf{y}}_\mathbf{i})^2 - c\sum_{j=1}^{p}\beta_j^2 \tag{3.4}$$

where n is the number of observations, $p-1$ is the number of predictor variables, $c \geq 0$ is a biasing constant, Y_i is the i_{th} observation and $\hat{y}_i$ is the prediction for Y_i given by (3.3). Therefore the modified least squares estimators are:

$$\mathbf{b^r} = (\mathbf{X'X} + c\mathbf{I})^{-1}\mathbf{X'Y} \tag{3.5}$$

Let $\mathbf{W} = (\mathbf{X'X} - c\mathbf{I})^{-1}$ hence,

$$\text{Bias}(\mathbf{b^r}) = -c\mathbf{W}\boldsymbol{\beta} \ ,$$

$$\text{Var}(\mathbf{b^r}) = \sigma^2\mathbf{W}\mathbf{X'}\mathbf{X}\mathbf{W} \ ,$$

$$\text{and E}(\mathbf{b^r}) = \mathbf{W}(\mathbf{X'X})\boldsymbol{\beta}$$

There always exists a value c, such that $\mathbf{b^r}$ has a smaller mean squared error than $\mathbf{b}$. When c increases, the bias in $\mathbf{b^r}$ will increase but the variance will decrease.

3.1.1 Ridge Trace

Let Y^* and X_{ik} defined by

$$Y^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \bar{Y}}{s_y}\right)$$

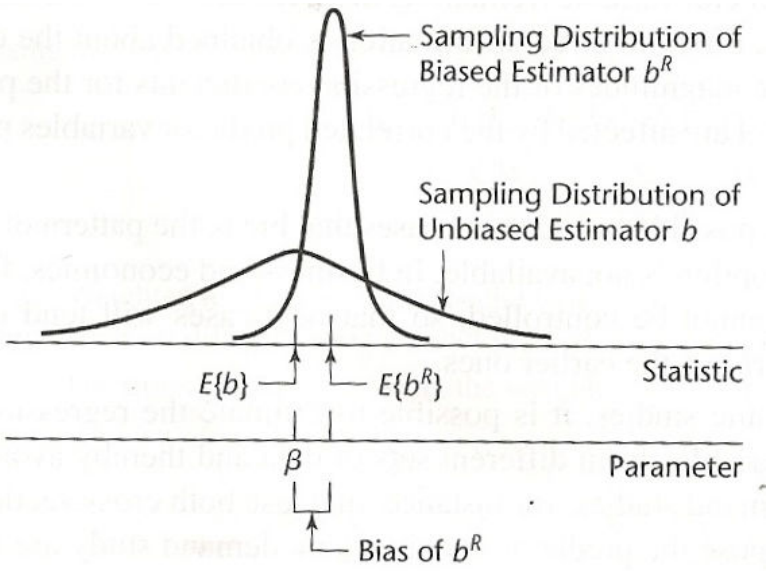


Figure 3.1: Sampling distribution of b. *Figure taken from Kutner et al. (2005).*

$$X_k^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_k - \bar{X_k}}{s_k}\right)$$

where $k = 1, ..., p-1$, $\bar{Y}$ and $\bar{X_k}$, are the respective means of the Y and the X_k observations, s_y and s_k are the respective sample standard deviation. The regression model in the variables Y^* and X_k^* is called a standarized regression model.

A method to determine the "optimum "value of c is the ridge trace. The values of the ridge standarized regression coefficients vary widely as c change slightly from 0. However, as c increases the fluctuation ceases and the regression coefficients tend to move slowly toward zero [13]. The essential idea is to plot the values of the $p-1$ ridge standarized regression coefficients for differents values of c, usually in the interval of $[0, 1]$, then pick the smallest value of c that produces a stable estimate of $\mathbf{b^r}$. See Figure 3.2

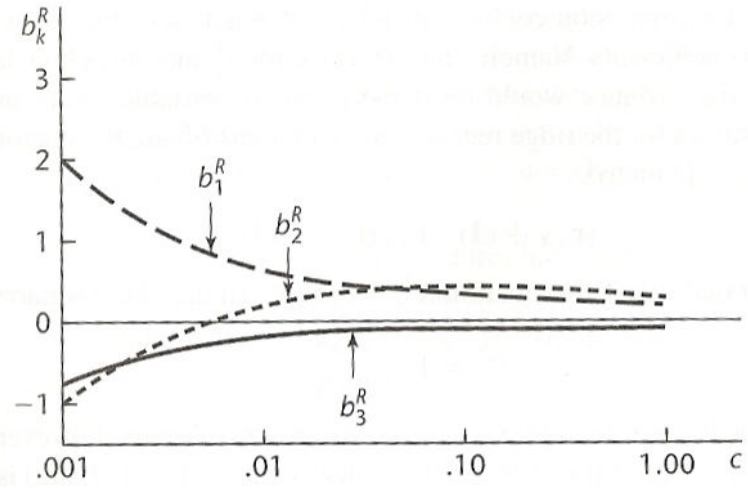


Figure 3.2: Ridge Trace. *Figure taken from Kutner et al. (2005).*

3.2 Nonparametric Regression Using P-splines

3.2.1 Nonparametric Regression

Suppose that Y is observed at n predetermined values of the independent variable x. Having observed x, the expected value of Y is given by the regression function. Let $f(x)$ be the regression function then the regression relationship between Y and x can be modeled as

$$y_i = f(x_i) + \epsilon_i \, i = 1, ..., n \tag{3.6}$$

where ϵ_i are zero mean, uncorrelated random errors with common variance σ^2 and $x_i \in [a, b]$ A parametric regression model assumes that $f(x)$ is known except for finitely many unknown parameters. A nonparametric regression model generally only assumes that $f(x)$ belongs to some infinite dimensional collection of functions [7].

3.2.2 Nonparametric Regression Using Splines

Splines (piecewise polynomials) provide a way of approximating nonlinear functions. The basic example of splines is the p_{th} degree trucated power function defined as

$$(x-t)_+^p = (x-t)^p I_{x>t}(x) \ , \tag{3.7}$$

where t is any real number. As a function of x this function takes on the value of 0 to the left of t and the values of $(x-t)^p$ to the right of t. In this case t is called a knot.

Splines of order r

The definition of a spline of order r is any function s of the form:

$$s(x) = \sum_{i=0}^{r-1} \theta_i x^i \ + \ \sum_{i=1}^{k} \delta_i (x-t_i)_+^r - 1 \tag{3.8}$$

for some set of real coefficients $\theta_0, ..., \theta_{r-1}, \ \delta_0, ..., \delta_k$. From the definition we can conclude that:

- s is a piecewise polynomial of order r on any subinterval $[t_i, t_{i+1})$.
- s has $r-2$ continuous derivatives.
- s has an $(r-1)$st derivative that is a step function with jumps at $t_1, ..., t_k$.

Let $S^r(t_1, ..., t_n)$ be the set of all splines of order r, $S^r(t_1, ..., t_n)$ is a vector space with dimension $n+r$. Note that $\mathbf{B} = \{1, t, t^2, ..., t^{r-1}, (t-t_1)_+^{r-1}, (t-t_2)_+^{r-1}, ..., (t-t_n)_+^{r-1}\}$ is a basis set for $S^r(t_1, ..., t_n)$.

Natural Splines of order r

A spline is called a natural spline if s, as defined in (3.8), is a polynomial of order m outside of $[t_1, t_n)$. Here $k = n$ and $t_i = x_i$. A natural spline satisfies [7]:

- $s^{(j)}(a) = s^{(j)}(b) = 0$, $j = m, ..., 2m-1$. Where $a \leq t_1$ and $b \geq t_n$.
- $\theta_m = ... = \theta_2 m - 1 = 0$ in (3.8).

Smoothing Splines

Suppose $f(.)$ as defined in (3.6) belongs to $W_2^m[a,b]$ were is defined as $W_2^m[a,b] = \{f|f^{(m)}$ is continuous on $[a,b]$ and square integrable$\}$. The natural measure of smoothness associated with $f(.)$ is $\int_a^b [f^{(m)}(x)]^2 dx$ and the standard measure of goodness-of-fit to data is $n^{-1}\sum_{j=1}^n (y_j - f(x_j))^2$. Then we try to minimize

$$n^{-1}\sum_{j=1}^{n}(y_j - f(x_j))^2 + \lambda \int_a^b [f^{(m)}(x)]^2 dt, \;\; \lambda \geq 0 \tag{3.9}$$

The parameter λ is called the smoothing parameter. The minimizer function is a natural spline of order $r = 2m$ with knots at $t_1, ..., t_k$ [7].

B-Splines

The basis set of functions $\mathbf{B}$ is called the truncated power basis. This basis makes the parameter estimation difficult to solve [7]. Another basis is the B-spline basis, this makes the estimation easier because we need fewer operations to get the estimates. Let $t_1, t_2, ...t_n$ be some given knots on the interval $[a,b]$. To develop the B-spline basis for $S^r(t_1, ...t_n)$, first we have to define $2r$ additional knots:

$$t_{-(r-1)} = ... = t_0 = a \quad \text{and} \quad t_{n+1} = ... = t_{n+r} = b \; .$$

The B-spline of order r corresponding to the knots $t_i, ..., t_{i+r}$ can be defined by the recursion formula,

$$N_{i,r}(t) = \frac{t - t_i}{t_{i+r-1} - t_i} N_{i,r-1}(t) + \frac{t_{i+r} - t}{t_{i+r} - t_{i+1}} N_{i+1,r-1}(t)$$

with

$$N_{i,1}(t) = \begin{cases} 1 & t \in [t_i, t_{i+1}] \\ 0 & \text{otherwise} \; . \end{cases}$$

Some properties of B-Splines [6] are:

- The B-spline of order r is positive on a domain spanned by $r+1$ knots, everywhere else it is zero.

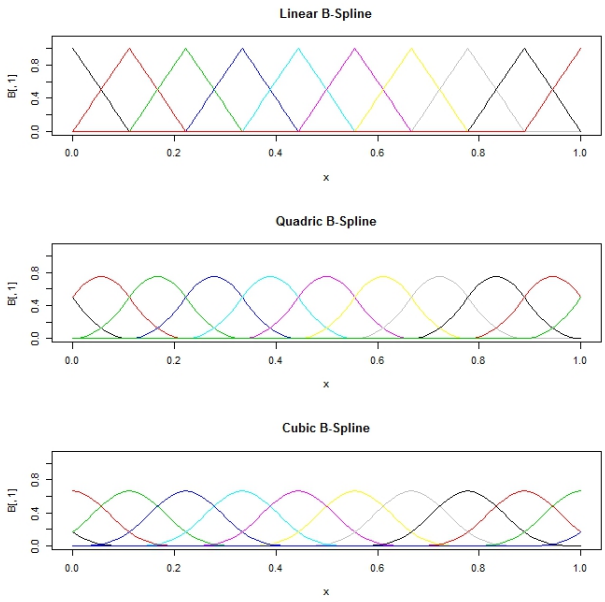


Figure 3.3: Examples of B-Splines, Linear (order 2), Quadratic (order 3) and Cubic (order 4).

- Except at the boundaries, it overlaps with $2(r-1)$ of its neighbors.
- At a given x, only r B-splines are nonzero.

To ilustrate these properties, Figure (3.3) shows examples or B-splines of order 2, 3 and 4 with 9 interior points. Now, the regression curve $f(.)$ as defined in (3.6) can be approximated by

$$f(t) = \sum_{j=1}^{n} \beta_j N_{j,r}(t) \ ,$$

where $N_{j,r}(t)$ is the value at t of the j^{th} B-spline of order r. And the penalized sums of squares are:

$$\sum_{i=1}^{n} \{y_i - \sum_{j=1}^{k} \beta_j N_{j,r}(t_i)\}^2 + \lambda \int_a^b \{\sum_{j=1}^{k} \beta_j N''_{j,r}(t)\}^2 dt \tag{3.10}$$

where n is the number of observations and k is the number of B-splines.

P-Splines

The penalty part of (3.10) $\int_a^b \{\sum_{j=1}^x \beta_j N''_{j,r}(t)\}^2 dt$ depends on the number of knots. A large number of knots makes the calculations difficult and with a small number of knots we underfit the data. Eilers and Marx (1996) avoid this problems by the use of P-Splines [6]. They proposed a penalty based on finite differences of coefficients of adjacent B-Splines

$$S = \sum_{i=1}^{n} \{y_i - \sum_{j=1}^{k} \beta_j N_{j,r}(t_i)\}^2 + \lambda \sum_{j=m+1}^{n} (\Delta^m \beta_j)^2$$

where,

$$\Delta\beta_j = \beta_j - \beta_{j-1}$$

$$\Delta^2\beta_j = \Delta(\Delta\beta_j) = \beta_j - 2\beta_{j-1} + \beta_{j-2} \text{ , etc.}$$

They show that the difference penalty retains the flavor of the integrated square of the m^{th} derivative; and the polynomial regression models occur as limits for large values of λ. The value of β minimizing equation (3.10) with the P-spline penalty is

$$\hat{\beta} = (B^T B + \lambda D_m^T D_m)^{-1} B^T y$$

where D_m is the matrix representation of the difference operator Δ_m, β is the vector of β_j's and B is the B-spline matrix with elements $b_{ij} = N_{j,r}(t_i)$.

For a penalty on the second-order differences of the B-spline coefficient,

$$\hat{\beta} = (B^T B + \lambda D_2^T D_2)^{-1} B^T y \tag{3.11}$$

where,

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & \cdots & 0 \\ 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 1 & -2 & 1 \end{bmatrix}_{(k-2)\times k} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}_{k\times 1}$$

Figure 3.4 shows the effect of the value of λ on the fitted values.

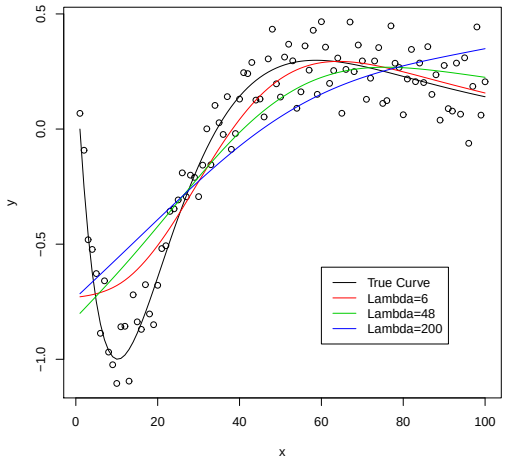


Figure 3.4: The effect of the smoothing parameter on P-splines

3.3 Using the Functional Linear Model to Explore the Choice of the Lag for the Pollution Model

Functional linear models are used when the independent variable of the model is a functional variable, when the response variable is a functional variable or both cases. In this thesis we focus on the case when just one or more of the covariates are functional variables. The model is:

$$Y = \alpha + \int_{t_0}^{t_1} x(s)\beta(s)ds + \epsilon, \tag{3.12}$$

where $x(s)$ is a covariate that is a function of the continuous variable $s \in [t_0, t_1]$. This model is the functional version of the simple linear regresion model $Y = X\beta + \epsilon$.

3.3.1 Historical Functional Linear Model

A functional linear model is called historical functional linear model if the response depends on past values of the independent variable. For example, suppose that $y_i(t)$ is an indicator of the hospitalization of the individual i at time t, $t = 1, 2, .., T$, and $x(s)$ is the covariate air pollutant NO_2 level at time s. It is not logical that the hospitalization depends on the

level of NO_2 at time t (e.g. the response of the body is instantaneous and does not depend on past exposures), in this case $x(s)$ is used as covariate in times $s \leq t$. Researchers limit s to be in $[t-\delta, t]$ for some lag δ. Let η_t be a function of $E(Y(t))$ such that the truncated historical functional model for η_t is:

$$\eta_t = \alpha(t) + \int_{s(0)}^{t} x(s)\beta(t-s)ds, \quad s(0) = max(0, t-\delta).$$

Suppose that s is hourly indexed and t is daily indexed. Define the function $S(t) = 24t$ to convert the daily index t in the response to the hourly index s of the covariate. By a change of variables, the historical functional model is:

$$\eta_t = \alpha(t) + \int_{0}^{\min(S(\delta),S(t))} x(S(t)-s)\beta(s)ds. \tag{3.13}$$

3.3.2 Exploring the Choice of the Lag for the Pollution Model

Preprocessing of the data

We explained before why it is not logical that the hopitalization on day t depends just on the pollution exposure at time t. Also, it is not logical assume that all the past pollution exposures affect the hospitalizations on day t. We truncated our historical linear model by defining a lag time δ. Yang(2005) explains how the truncation affects the estimate of $\beta(.)$. Thus, the truncated historical function linear model for η_t is

$$\eta_t = \alpha(t) + \int_{0}^{min(24t,24\delta)} x(24t-s)\beta(s)ds \ . \tag{3.14}$$

Let the slope function $\beta(.)$ be approximated by a linear combination of B-splines or order r:

$$\beta(s) = \sum_{j=1}^{K} b_j N_{j,r}(s)$$

with K equal to the number of B-splines (number of interior knots + r). Thus,

$$\begin{aligned}\eta_t = \int_0^{min(24t,24\tau)} x(24t-s)\beta(s)ds &= \int_0^{min(24t,24\delta)} x(24t-s)\Big\{\sum_{j=1}^{K} b_j N_{j,r}(s)\Big\}ds \\ &= \sum_{j=1}^{K} b_j\Big\{\int_0^{min(24t,24\delta)} N_{j,r}(s)x(24t-s)ds\Big\} \\ &= \sum_{j=1}^{K} b_j\phi_j(t;x). \qquad (3.15)\end{aligned}$$

Here $\phi_j(t;x) = \int_0^{min(24t,24\delta)} N_{j,r}(s)x_+(24t-s)ds$ and

$$x_+(24t-s) = \begin{cases} x(24t-s) & \text{when } s \le \min(24t, 24\delta) \\ 0 & \text{otherwise} \end{cases}$$

for $s \in [0, 24T]$ and $t \in \{1, ..., T\}$. The next step is approximate the integral in $\phi_j(t;x)$ with a quadrature rule using Q equally spaced time points $\{g_q\}_{q=1}^{Q}$ in [0,24T]:

$$\phi_j(t;x) \approx \frac{24T}{Q}\sum_{q=1}^{Q} x_+(24t-g_q)N_{j,r}(g_q)$$

Let N denote a $Q \times K$ matrix with $N_{i,j} = N_{j,r}(g_i)$, $X^t(t) = (x_+(24t-g_1), ..., x_+(24t-g_Q))$, $\frac{24T}{Q} = 1$ and

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}_{K\times 1}.$$

Then the matrix approximation to (3.15) is

$$\zeta(t) = X^t(t)Nb.$$

The estimation of $\alpha(.)$ is ignored in conditional logistic regression models (see details in

Chapter 2). Computing $X^t(t)N \approx \phi_j(t;x)$ is called the preprocessing step [4]. To estimate the coefficients in equation (3.15), we use the ridge regression function COXPH in R with a modified penalty for ridge regression in the conditional likelihood. The covariates in that ridge regression are $\phi_1(t;x), ..., \phi_k(t;x)$. The smoothing parameter λ as in equation (3.11) is chosen with a modified ridge trace method.

Choice of the smoothing parameter

In our analysis $\delta = 8$ (i.e. 8 days before the day of hospitalization). After preprocessing the NO_2 and $PM_{2.5}$ data, we need to calculate the B-spline coefficients for the cubic P-splines. Recall that the estimation of β

$$\hat{\boldsymbol{\beta}} = (\mathbf{B^T B} + \lambda \mathbf{D_2^T D}_2)^{-1} \mathbf{B}^T \mathbf{y}$$

as defined in (3.11), is similar to the estimation of β in ridge regression

$$\mathbf{b^r} = (\mathbf{X'X} + c\mathbf{I})^{-1} \mathbf{X'Y}$$

as defined in (3.5). The COXPH function in R fits a conditional regression model and we can specify the ridge regression terms in the model. The ridge penalty part is modified to obtain a P-splines [4]. The semiparametric model used is

$$\begin{aligned} \text{logit}[Pr(Y(t)=1)] &= \alpha(t) + AT_{t-1} + \\ &\quad \beta_1(\text{low wind}_{t-3}) + \beta_2(\text{high wind}_t) + \\ &\quad \int_0^{min(24t,24\delta)} x(24t-s)\beta_x(s)ds + \int_0^{min(24t,24\delta)} z(24t-s)\beta_z(s)ds. \end{aligned} \tag{3.16}$$

where x represents the data for NO_2 and z represents the data for $PM_{2.5}$, AT_{t-1} represents the apparent temperature of the previous day which is included in the model as a piecewise linear function with two equally spaced interior knots without penalty, the calculation of the AT is explained in Chapter 4. In the ridge trace method we plot the values of the

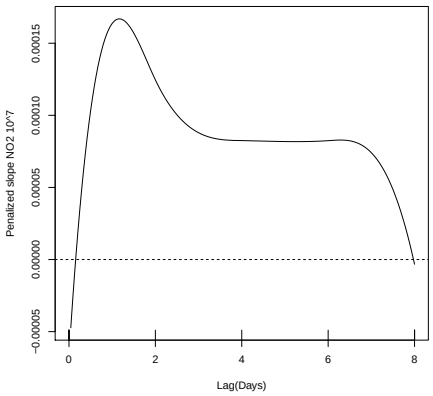
ridge regression coefficients that were standarized by dividing by their standard error, we did it for different values of c as in Bulathsinhala (2011), usually in the interval of $[0, 1]$. Let $c = \frac{\lambda}{\lambda_{max}}$, $0 \leq \lambda \leq \lambda_{max}$, λ_{max} is the value of the smoothing parameter when the slope function becomes a straight line.

Asthma

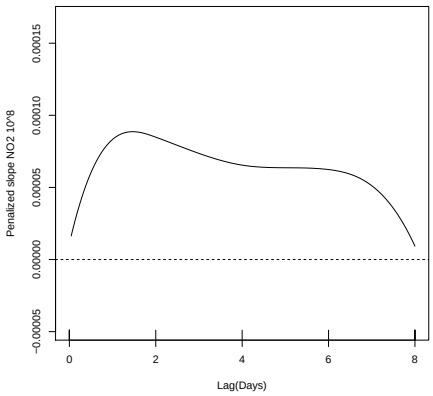
When $[Pr(Y(t) = 1)]$ in (3.16) represents the probability fo being hospitalized due to asthma, the value of λ_{max} is equal to 10^{10} for both NO_2 and $PM_{2.5}$, see Figure 3.5 and Figure 3.6.

Ridge trace plots for NO_2 and $PM_{2.5}$ were obtained by plotting $\hat{b_j}/SE(b_j)$ (standardized coefficients) versus the values of c in table 3.1, see Figure (3.7) and Figure (3.8). We can see that the standardized coefficients start to stabilize around $c = 0.4$ in the NO_2 and $PM_{2.5}$ plots. The penalized conditional logistic regression model in (3.16) was fitted with $\lambda = 4 \times 10^9$ for NO_2 and $PM_{2.5}$. The fit using P-splines is presented in Figure (3.9). From Figure (3.9) we cannot choose the lags because the curves are over smoothed. Finally, we opted to choose the greatest value for λ which allows us to see additional features of the regression coefficient. The value chosen is $\lambda = 10^8$, a comparison between $\lambda = 4 \times 10^9$ and $\lambda = 10^8$ is in Figure 3.10.

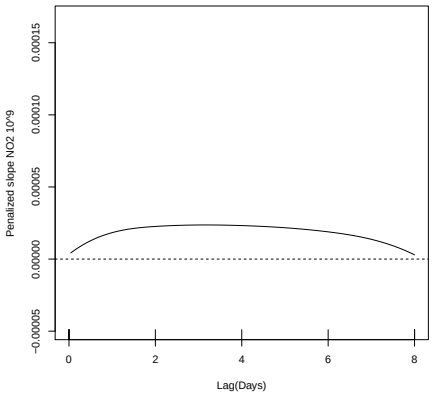
The penalized conditional logistic regression model in (3.16) was fitted with $\lambda = 10^8$ for NO_2 and $PM_{2.5}$ B-splines coefficients the result is presented in Figure (3.11), and by visual inspection the average of lags 0-2 for NO_2 and average of lags 0-1 for $PM_{2.5}$ was chosen for the parametric regression models. Until this step we have worked simultaneously with NO_2 and $PM_{2.5}$, we ran the single pollutant models to verify that if we work with the single pollutant models the pollution effects are the same, from Figure(3.12) we can see that the conclusions similar. Finally, for wind speed lags we use lag zero for high wind speed and lag 3 for low wind speed as in Grineski et al.(2011) and Bulathsinhala (2011). The relationship between the coefficients of the B-splines functions and the coefficients of the chosen lags in the parametric model can be understood using Figure (3.13), here the solid line is the unpenalized slope function for $PM_{2.5}$, and the dotted line is the value of the coefficients



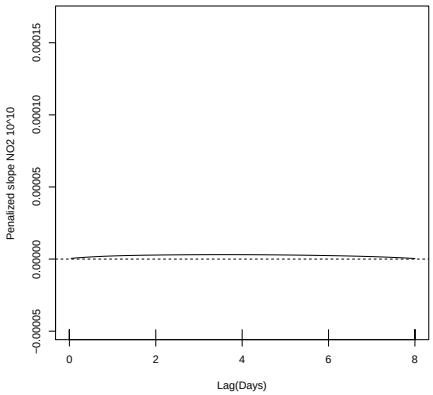
(a) NO_2 Slope function when $\lambda = 10^7$



(b) NO_2 Slope function when $\lambda = 10^8$

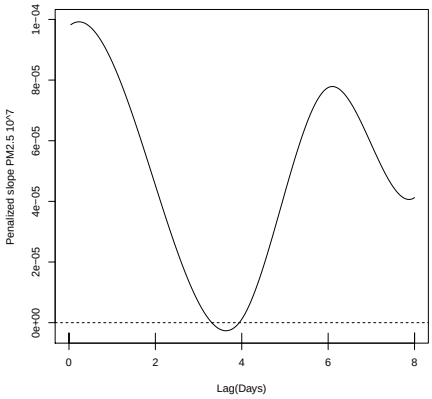


(c) NO_2 Slope function when $\lambda = 10^9$

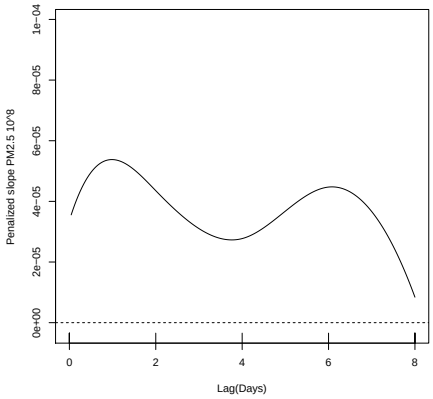


(d) NO_2 Slope function when $\lambda = 10^{10}$

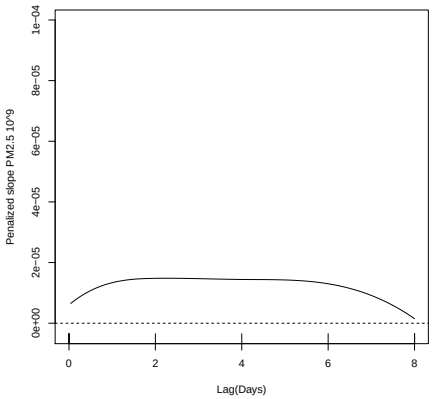
Figure 3.5: The slope function of NO_2 at different values of the smoothing parameter λ.



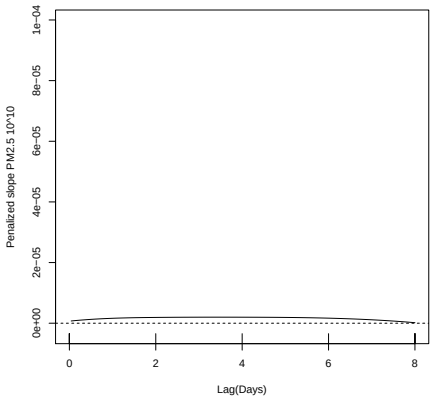
(a) $PM_{2.5}$ Slope function when $\lambda = 10^7$



(b) $PM_{2.5}$ Slope function when $\lambda = 10^8$



(c) $PM_{2.5}$ Slope function when $\lambda = 10^9$



(d) $PM_{2.5}$ Slope function when $\lambda = 10^{10}$

Figure 3.6: The slope function of $PM_{2.5}$ at different values of the smoothing parameter λ.

Table 3.1: Corresponding λ values for c.

λ	c
0	0
10^7	0.001
3×10^7	0.003
5×10^7	0.005
7×10^7	0.007
9×10^7	0.009
10^8	0.01
3×10^8	0.03
5×10^8	0.05
7×10^8	0.07
10^9	0.1
1.5×10^9	0.15
2×10^9	0.2
3×10^9	0.3
4×10^9	0.4
5×10^9	0.5
6×10^9	0.6
7×10^9	0.7
8×10^9	0.8
9×10^9	0.9
10^{10}	1

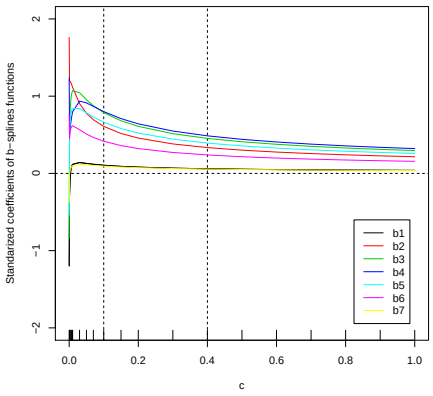


Figure 3.7: NO_2 Adapted Ridge Trace with $\lambda_{max} = 10^{10}$

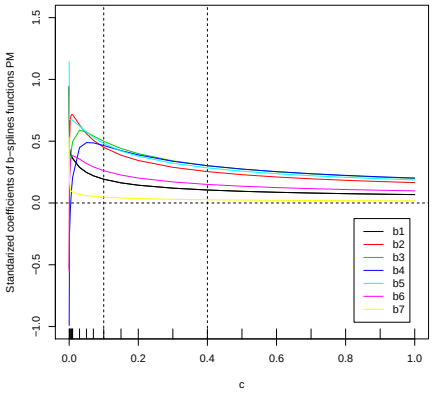
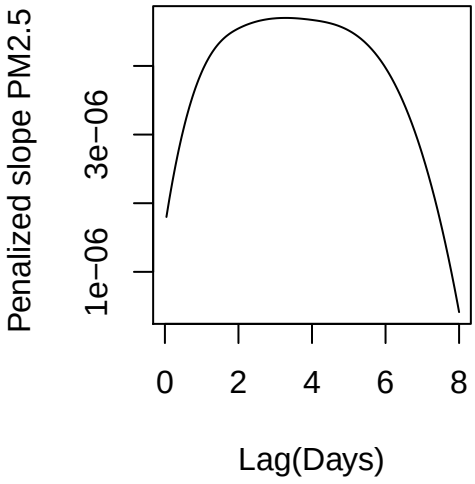


Figure 3.8: $PM_{2.5}$ Adapted Ridge Trace with $\lambda_{max} = 10^{10}$



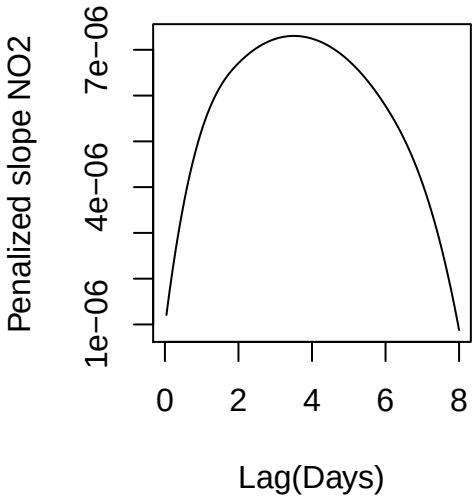
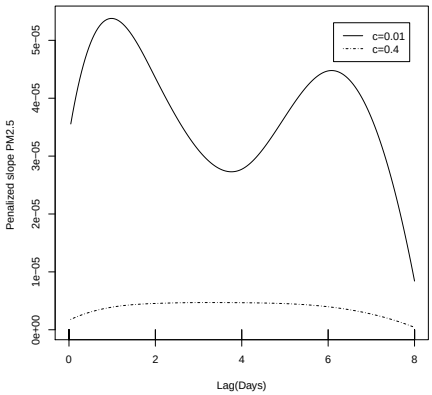
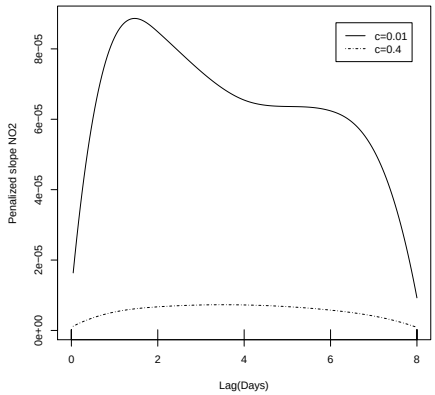


Figure 3.9: Slope function of β with $\lambda = 4 \times 10^9$



(a) $PM_{2.5}$ with $\lambda_{max} = 10^{10}$



(b) NO_2 with $\lambda_{max} = 10^{10}$

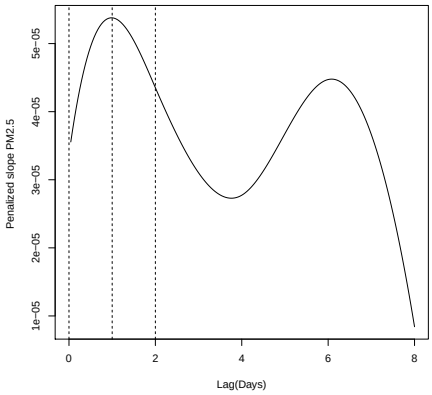
Figure 3.10: Comparison between two levels of smoothing.

in the parametric model of the lags covered by that interval divided by the the number of hours in the interval, for example, the value of the dotted line between lag 0 and 1 is equal to β_3 in the parametric model (3.17) divided by 48.

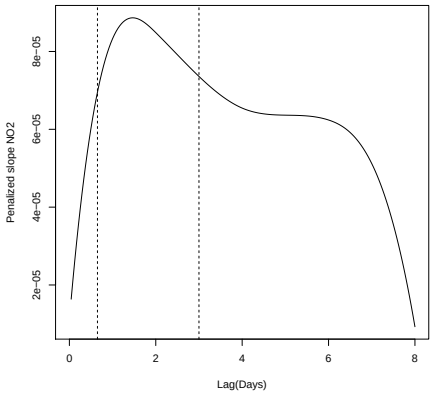
$$\begin{aligned} \text{logit}[Pr(Y(t)=1)] &= \alpha(t) + AT_{t-1} + \\ &\quad \beta_1(\text{low wind}_{t-3}) + \beta_2(\text{high wind}_t) + \\ &\quad \beta_3(\text{avg. PM }_{2.5} \text{ lags 0-1}) + \beta_4(\text{avg. PM}_{2.5} \text{ lags 2-4}) + \\ &\quad \beta_5(\text{avg. PM}_{2.5} \text{ lags 5-7}). \end{aligned} \tag{3.17}$$

Chronic Obstructive Pulmonary Disease (COPD)

The lag for low and high wind speed for the COPD analysis (see equation 3.18) was selected using the same procedure as for asthma. The smoothing parameter λ_{max} for NO_2 is equal to 10^{10}, and for $PM_{2.5}$ λ_{max} is equal to 10^9. From Figures (3.14) and (3.15) the final values of the smoothing parameter λ are 2×10^9 for NO_2 and 4×10^8 for $PM_2.5$.

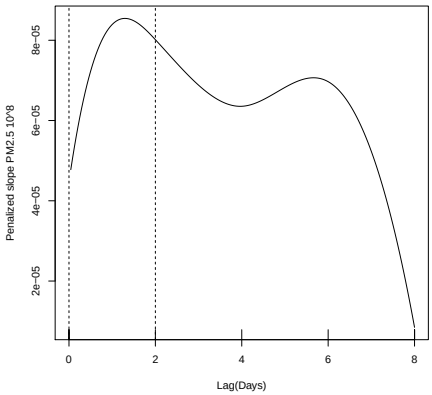


(a) $PM_{2.5}$ with $\lambda = 10^8$

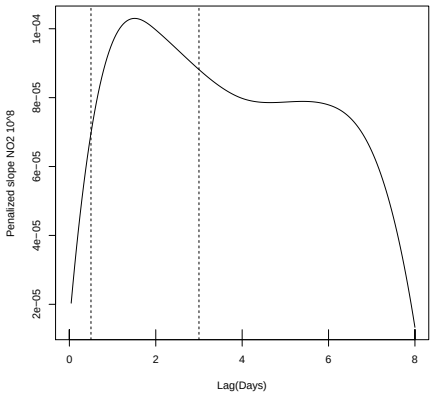


(b) NO_2 with $\lambda = 10^8$

Figure 3.11: Final choice of Lags.



(a) Single pollutant model result for $PM_{2.5}$ with $\lambda = 10^8$



(b) Single pollutant model result for NO_2 with $\lambda = 10^8$

Figure 3.12: Single pollutant models.

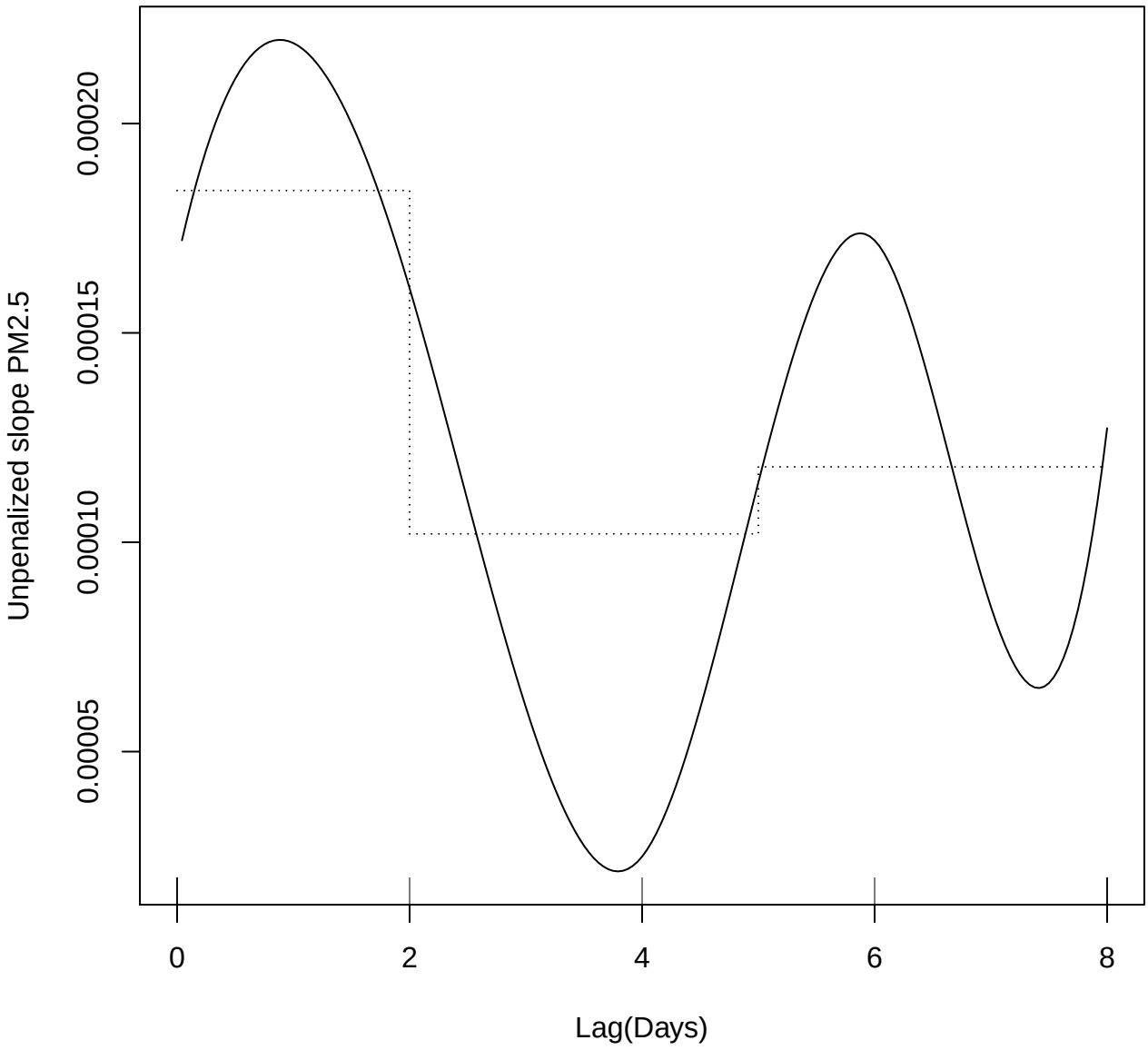


Figure 3.13: The relationship between the coefficients of the B-splines functions for $PM_{2.5}$ and the coefficients in the parametric model.

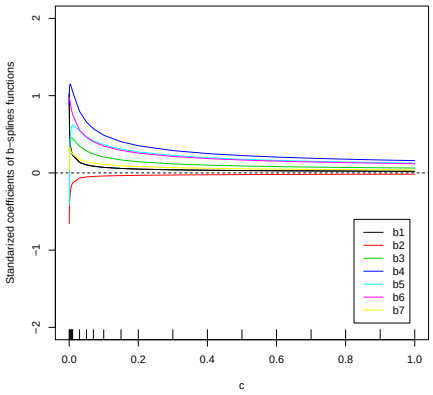


Figure 3.14: NO_2 Ridge Trace with $\lambda = 10^{10}$. The coefficient estimates are stabilized around $c = 0.2$.

$$\begin{aligned} \text{logit}[Pr(Y(t)=1)] &= \alpha(t) + AT_{t-1} + \\ &\beta_1(\text{low wind}_{t-3}) + \beta_2(\text{high}_{t-3}) + \\ &\int_0^{min(24t,24\delta)} x(24t-s)\beta_x(s)ds + \int_0^{min(24t,24\delta)} z(24t-s)\beta_z(s)ds. \end{aligned} \tag{3.18}$$

where x, z and AT_{t-1} defined as in (3.16) and $Pr(Y(t) = 1)$ equal to the probability of being hospitalized due to a COPD.

For COPD we do not have any previous work for wind speed so that we plot the slope functions of the B-splines for NO_2 and $PM_2.5$ with lag zero to lag 3 for wind speed. From Figures (3.16) and (3.17) we chose the average of lags 2-6 for the parametric model with NO_2 and lag 6 for the parametric model with $PM_{2.5}$ and conclude that the models are not sensitive to the chosen lags for wind speed. Single pollutants model are shown in Figure (3.18), we can see that the we obtain the same results with single pollutants models. Once we chose the lags for the pollutants, we ran the parametric conditional logistic regresion

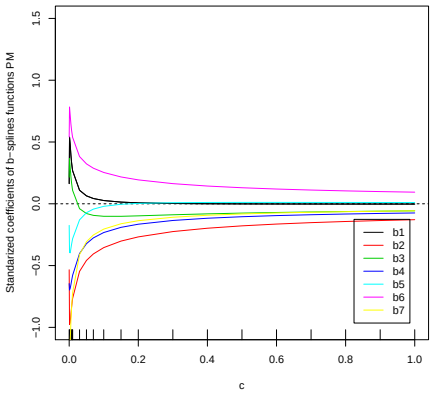


Figure 3.15: $PM_{2.5}$ Ridge Trace with $\lambda = 10^9$. The coefficient estimates are stabilized around $c = 0.4$

model with wind speed lags equal to 0, 1, 2 and 3. The odds ratio and the 95% confidence interval for each lag was plotted in Figures (3.19) and (3.20). Based on those plots, lag zero for low and high speed for both pollutants was chosen because these had the highest odd-ratio.

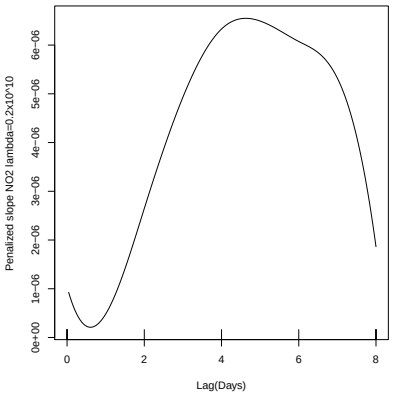
3.4 Interpretation of the Coefficients in Conditional Linear Regression

Odds ratio

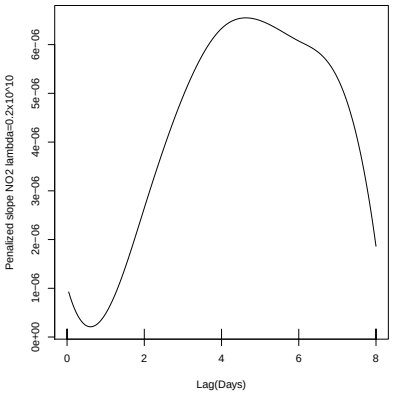
The odds ratio is defined as

$$\text{odds}|X = \frac{\text{P}(Y = 1|X)}{\text{P}(Y = 0|X)},$$

where X represents the covariates and Y is the response variable. The odds ratio for the covariates in status A $(X = X_A)$ versus the covariates in status B $(X = X_B)$ is equal to $\frac{\exp\{\beta X_A\}}{\exp\{\beta X_B\}} = \exp\{\beta X_A - \beta X_B\} \approx \frac{\text{P}(Y=1|X_A)}{\text{P}(Y=1|X_B)}$, that means that we supposed that the event is rare enough to consider $\text{P}(Y = 0|X)$ close to 1. For rare events, the relative risk is an approximation for the odd ratio [25].

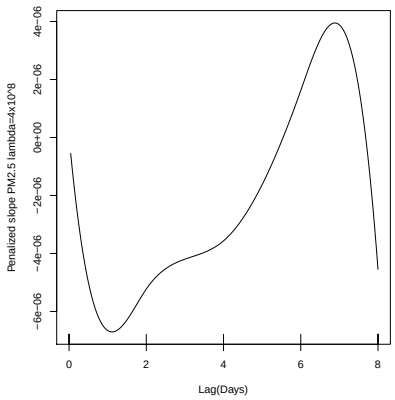


(a) The slope function for NO_2 with $\lambda = 2 \times 10^9$ and lag zero for low and high wind speed.

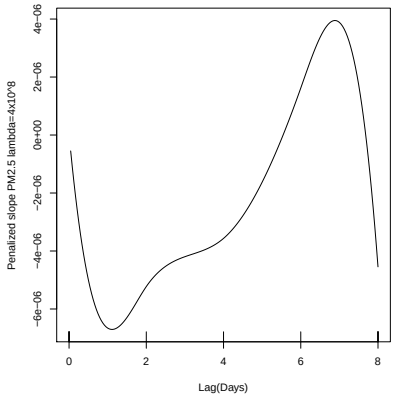


(b) The slope function for NO_2 with $\lambda = 2 \times 10^9$ and lag 3 for low and high wind speed.

Figure 3.16: The shape of the slope functions for NO_2 does not seem be sensitive to the lag chosen for wind speed.

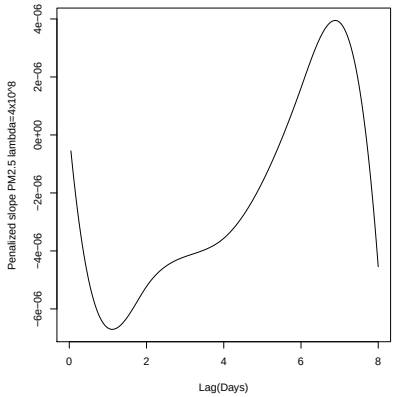


(a) The slope function for $PM_{2.5}$ with $\lambda = 4 \times 10^8$ and lag zero for low and high wind speed.

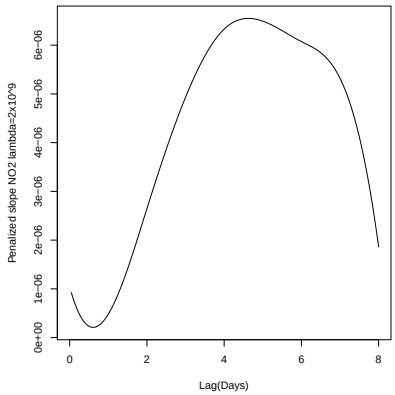


(b) The slope function for $PM_{2.5}$ with $\lambda = 4 \times 10^8$ and lag 3 for low and high wind speed.

Figure 3.17: The shape of the slope functions for $PM_{2.5}$ does not seem be sensitive to the lag chosen for wind speed.



(a) Single pollutant model result for $PM_{2.5}$ with $\lambda = 4 \times 10^8$



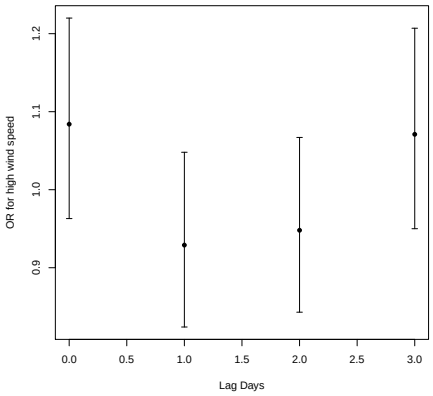
(b) Single pollutant model result for NO_2 with $\lambda = 2 \times 10^9$

Figure 3.18: Single Pollutants models.

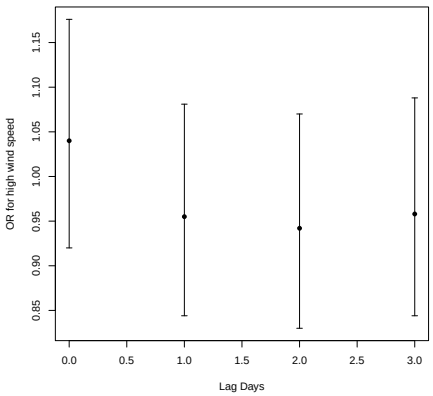
Examples of the Interpretation of the Coefficients When we use interactions in the conditional logistic regression model, the interpretation of the estimators of β of the categorical variables are not directly the exponential of the odds ratio of $x_i = 1$ vs the reference. Suppose we have the model

$$\begin{aligned} \text{logit}[Pr(Y(t)=1)] &= \alpha(t) + \beta_1(PM2.5) + \beta_2(NO_2) \\ &\quad + \beta_3(A * NO_2) + \beta_4(A * NO_2) + \beta_5(E * NO_2) + \beta_6(Y * NO_2) \\ &\quad + \beta_7(W * NO_2) + \beta_7(M * NO_2). \end{aligned} \tag{3.19}$$

where the variables were designed as in Table 3.2. If we want to calculate the odds ratio when $X_A =$ Female Hisp. Adults at $\text{NO}_2 = c$ vs $X_B =$ Female Hisp. Adults at $\text{NO}_2 = c+10$ we need to write the model for X_A then we write the model for X_B, finally we calculate the difference between the two models and the result is the log of the odds ratio, see Table 4.14. Consider another example, if we want to calculate the odds ratio between $X_A =$ Adults versus $X_B =$Children, as before, we write the model for X_A and the model for X_B,

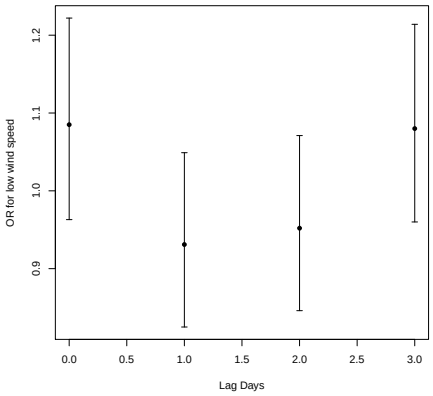


(a) Selection of lag of low wind speed for use in the single model pollution with NO_2

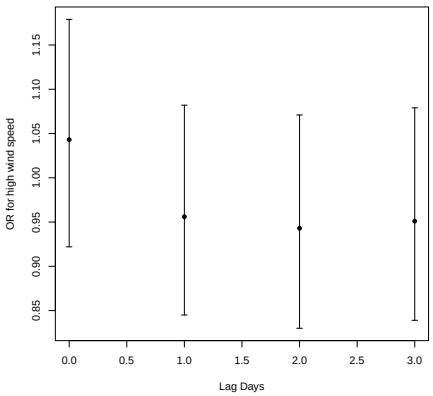


(b) Selection of lag of high wind speed for use in the single model pollution with NO_2

Figure 3.19: Selection of wind speed lags for COPD model. 95% confidence intervals are plotted



(a) Selection of lag of low wind speed for use in the single model pollution with $PM_{2.5}$



(b) Selection of lag of high wind speed for use in the single model pollution with $PM_{2.5}$

Figure 3.20: Selection of wind speed lags for COPD model. 95% confidence intervals are plotted

then calculate the difference, in Table 3.4 we can see that the result depends on c that is the value of continuous variable NO_2, thus we need to define that value in order to get the odds ratio.

Table 3.2: Categorical variables

Age		Sex		Ethinicity	
Children	Ref	Female	Ref	Hispanics	Ref
Adults	A	Male	M	Non-Hisp Other	O
Elderly	E			Non-Hisp White	W
Young Adults	YA				

Table 3.3: Female Hisp. Adults $NO_2 = x$ vs Female Hisp. Adults $NO_2 = x + 10$

	Estimator of Coefficients							
Terms in the Odds Ratio:	$PM_2.5$	NO_2	A*NO_2	E*NO_2	YA*NO_2	O*NO_2	W*NO_2	M*NO_2
when NO_2 level increases by 10ppb	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
X_A =Female Hispanic Adults at NO_2 level=c+10	-	$\beta_2(c+10)$	$\beta_3(c+10)$	0	0	0	0	0
X_B =Female Hispanic Adults at NO_2 level=c	-	$\beta_2(c)$	$\beta_3(c)$	0	0	0	0	0
$X_A - X_B$	0	$\beta_2 \times 10$	$\beta_3 \times 10$	0	0	0	0	0

Table 3.4: Adults vs Children

	Estimator of Coefficients							
Terms in the Odds Ratio:	$PM_2.5$	NO_2	A*NO_2	E*NO_2	YA*NO_2	O*NO_2	W*NO_2	M*NO_2
	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
X_A =Adults	-	$\beta_2(c)$	$\beta_3(c)$	0	0	-	-	-
X_B =Children	-	$\beta_2(c)$	0	0	0	-	-	-
$X_A - X_B$	0	0	$\beta_3(c)$	0	0	0	0	0

Chapter 4

Statistical Analysis of the Data

4.1 Description of the Data

4.1.1 Hospitalization Data

The hospitalization data was obtained from the Texas Health Care Information Council (THCIC) in Austin, Texas. The fields that were used are: Record ID (unique number for each patient), Sex, Age, Race, Ethnicity, County, Diagnostic, Admision Date and Payer. We extracted the patients where the county was El Paso. We combined race and ethnicity into one variable as in Table 4.1, Non-Hispanic Others are mostly African Americans. The data is for the time period $2005 - 2010$.

Table 4.1: Creation of new variable that combines ethnicity and race information

Ethnicity	Race	ETH
Hispanic	Any	Hispanic
Non-Hispanic	White	Non-Hispanic White
Non-Hispanic	Non-White	Non-Hispanic Other

COPD

For COPD we extracted the patients with diagnostic codes $496.X$, $491.X$, and $492.X$ (ICD-9 codes) as in Peel et al. (2005) and Qiu et al. (2012). Patients less or equal to 44 years old were excluded because most of the time COPD is diagnosed in middle-aged or older Adults [18]. We included patients with Private Insurance, Medicare, Medicaid and Uninsured;

patients with Military Insurance were excluded because we do not have enough counts. After that there were 3702 hospitalizations due to COPD for the time period $2005-2010$.

Asthma

For asthma we extracted the patients with diagnostic code $493.X$ (ICD-9 code), patients less than 2 years old were excluded from the study because it is difficult to diagnose asthma in babies. We included patients with Private Insurance, Medicare, Medicaid, Military Insurance and Uninsured, those categories are used as surrogates for the SES of the patient. There were a total of 4953 addmissions due to asthma for the time period $2005-2010$.

The counts for COPD and asthma are in Tables 4.2 and 4.3. The probability of hospitalization (for either asthma or COPD) is less than one percent, hence the odds-ratio will be close to RR [39].

Table 4.2: Counts for COPD

COPD			
Category	Frequency	Category	Frequency
Adults (45-74yrs)	1634	Medicaid	185
Elderly (75yrs +)	2068	Medicare	2989
		Private	364
Hispanic	2179	Uninsured	164
Non-Hisp White	1204		
Non-Hisp Other	319		

4.1.2 Weather and Pollution Data

The weather hourly data source is the US National Weather Service, the data were collected at the El Paso International Airport for the period $2005-2010$. We extracted temperature, dew point and wind speed. We combined temperature and dew point in the apparent

Table 4.3: Counts for Asthma

Asthma			
Category	Frequency	Category	Frequency
Children (3-17yrs)	2161	Medicaid	1567
Young Adults (18-49yrs)	825	Medicare	1485
Adults (50-74yrs)	1257	Military	107
Elderly (75yrs+)	710	Private	1308
		Uninsured	486
Hispanic	3854		
Non-Hisp White	600	Female	2817
Non-Hisp Other	499	Male	2136

temperature variable defined as [29]

$$\text{Apparent Temperature} = -2.653 + 0.994\ \text{Temperature}(^0C) + 0.0153\ \text{Dew Point}^2(^0C).$$

The wind speed daily average was coded as in Table 4.4 in order to use wind speed as a categorical variable.

Table 4.4: Wind Speed Code. 2 m/s is the 10th percentile and 6 m/s is the 90th percentile of the daily average

Wind Speed (WS)	Code
WS $\leq$ 2 m/s	Low wind speed
2 $m/s <$ WS $\leq$ 6 m/s	Medium wind speed
6 $m/s <$ WS	High wind speed

Pollution hourly data was obtained from the Texas Commision of Evironmental Quality

Table 4.5: (TCEQ) monitoring stations.

CAMS
CAMS 12 (UTEP),
CAMS 37(Ascarate),
CAMS 40 (Sun Metro),
CAMS 41 (Chamizal),
CAMS 49 (Socorro), and
CAMS 72 (Skyline Park).

(TCEQ) website for the period $2005-2010$. There are six stations monitoring air pollutants in El Paso, see Table 4.5, we included data from CAMS 12, 37 and 41 for NO_2; data from from CAMS 12 and 40 for $PM_{2.5}$; data from CAMS 12, 37, 41, 49 and 72 for ozone, because only these had complete data.

Weather and pollution missing data values were imputed as follows for each CAMS site:

- Five or more consecutive missing hourly levels in a given day were replaced by the the mean of the available hourly observations of that day.
- Days with all hourly measurements missing were replaced with the mean of all available hourly over the six year period.
- B- spline interpolation was used to impute any missing observations spanning fewer than 5 hours.
- Negative values were replaced by zero.

Examples of imputation for the three cases are shown in Figures (4.1), (4.2) and (4.3) for NO_2 values from CAMS 12 . The summary statistics and correlation matrix for weather and pollution data are in Tables 4.6 and 4.7.

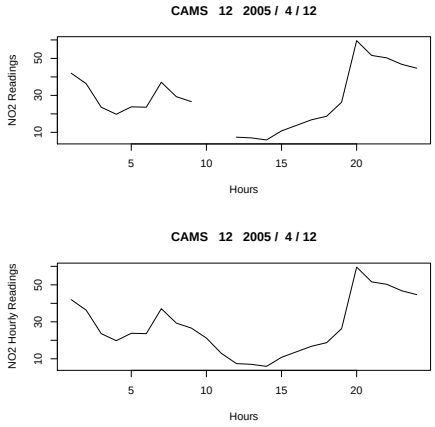


Figure 4.1: Fewer than five missing values, a B-spline interpolation was used to impute missing observations

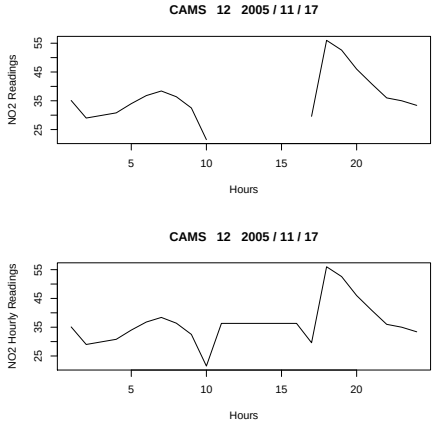


Figure 4.2: More than five, fewer than 24 missing values, the mean of the available observation of that day was used to impute missing observations

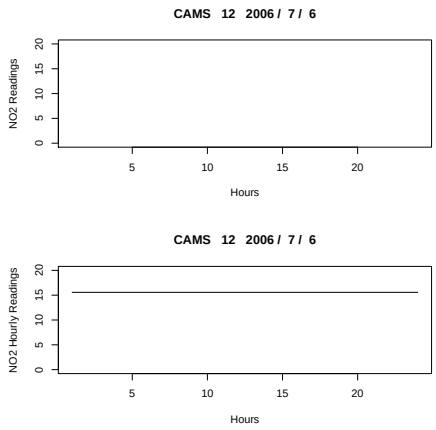


Figure 4.3: All hourly measurements are missing, the mean of all available hourly data over the six year period was used to impute missing observations

Table 4.6: Summary statistics for weather and pollution data

Variable	N	Min	10th Pctl	25th Pctl	Median	Mean	75th Pctl	90th Pctl	Max	SD
App. Temp. (^0C)*	2191	-3.33	5.67	10.28	17.89	17.58	25.69	28.53	32.36	8.64
Wind Speed (m/s)*	2191	0.41	1.97	2.55	3.37	3.76	4.51	6.02	13.06	1.78
Temperature (^0C)*	2191	-2.96	7.41	11.97	19.77	18.93	25.97	29.44	34.58	8.29
Dew Point (^0C)*	2191	-21.16	-10.83	-6.11	-0.07	1.13	9.49	14.24	18.77	9.25
$PM_{2.5}(\mu g/m^3)$**	2191	2.21	6.04	7.86	10.35	11.41	13.53	18.28	62.04	5.28
NO2(ppb)***	2191	5.37	17.67	24.91	33.67	33.86	41.97	49.67	91.73	12.35
O_3(ppm)****	2191	0.005	0.026	0.033	0.041	0.041	0.050	0.057	0.081	0.012

*Daily average.

**Daily average over the CAMS.

***Average over the CAMS of daily maximum. The EPA website uses 1-hour daily maximum of NO2 concentrations as quality the standard for air.

****Average over CAMS of daily 8 hrs. maximum. The EPA website uses average of 8-hour daily maximum O_3 concentrations as the quality standard for air.

	Wind Speed	$PM_{2.5}$	NO_2	O_3
Apparent Temp.	0.048	-0.093	-0.309	0.634
Wind Speed		-0.135	-0.402	0.087
$PM_{2.5}$			0.521	-0.032
NO_2				-.032

Table 4.7: Correlation Coefficients, N=2191

4.2 Methods and Results

The case-crossover design with analysis by conditional logistic regression (see Chapter 2) was used to determine if age, sex, race and SES modify the effects of daily air pollution on respiratory hospitalizations for the population in El Paso. For each case (patient hospitalized) we create a referent set of controls matching on the day of the week, in the same month and year as the event day. The covariates in our models are the pollutants as defined in Table 4.6 and their interactions with categorical variables defined in Table 4.8. Apparent temperature is included in the model as in equation (3.16). We work with single pollutant models because of the large correlation of 0.52 between NO_2 and $PM_{2.5}$, see Table 4.7.

Table 4.8: Design of the Categorical Variables

Age		Sex		Ethnicity		Payer	
Children	Ref(Asthma)	Female	Ref	Hispanic	Ref	Private	Ref
Adults	A Ref(COPD)	Male	M	Non-Hisp Other	O	Medicaid	Mdc
Elderly	E			Non-Hisp White	W	Medicare	Mdr
Young Adults	YA					Uninsured	Un

4.2.1 Results

We explore an analysis using the complete set of admissions for each disease and some subgroup analyses by ETHNICITY for COPD and asthma.

For COPD the covariate Sex is not included because we do not have enough counts in some payer categories; most of the cases are under Payer Medicare. The relative risk for interactions are not calculated in the parameter estimate tables, because the parameter estimate of the interaction of a categorical covariate with a continuous covariate is not the log of the relative risk (see Chapter 3).

COPD Results

The lag chosen for $PM_{2.5}$ is lag 6, whereas for NO_2 it is the average of lags 3 to 6, both low wind speed and high wind speed are lag zero (see Chapter 3). The complete and subgroup

exploratory analysis results for weather variables are in Table A.9, showing that wind speed is significant for Non-Hispanic White in both NO_2 and $PM_{2.5}$ single pollutant models. AT, low wind speed and high wind speed have a differential effect across subgroups, so the complete model is less realiable. The results for the interaction and pollutant parameters are in Table A.11: the interaction of Non-Hispanic Other with NO_2 is positive and significant at $\alpha = 0.10$ and the interaction of Medicaid with NO_2 is significant for Non-Hisp Other subgroup $\alpha = 0.05$.

From Table 4.9 we can see that for an increment of 6 $\mu g/m^3$ (IQR for $PM_{2.5}$), there is not a significant relative risk in any category, however, we can see that in all categories the relative risks are higher for Elderly than for Adults, Medicaid has highest relative risks among payers and Private the lowest, Hispanic has the highest relative risks among ethnicities and Non-Hispanic Other has the lowest. These results can be seen in Figure (4.4) as well, here the relative risk and 95% confidence limits are plotted at the $PM_{2.5}$ value of 26.34 $\mu g/m^3$ (98% percentile of the $PM_{2.5}$ lag 6 values for cases and controls during the period $2005 - 2010$). In the same Table 4.9 we can see that contrary to what we saw with $PM_{2.5}$, in NO_2 the Hispanic has the lowest relative risks among ethnicities and the Adults relative risk are higher than Elderly's when NO_2 increases by 17 ppb (interquartile range for NO_2 during the period 2005-2011); and for Non-Hispanic White and Non-Hispanic Other that have No Insurance the relative risks were significant. The relative risk and 95% confidence limits were plotted in Figure (4.5) when the NO_2 level is equal to 53.34 ppb (98% percentile of the average of lag 3 to 6 for NO_2 values for cases and controls during the period $2005-2010$), here the relative risk of Non-Hispanic Others compared with Hispanics is greater than one and significant.

COPD Subgroup Analysis

From the results for the exploratory subgroup analyses by race in Table 4.10 , we can see that the relative risk for Medicare in the Hispanic subgroup is 1.049, meaning that for an increment of $6\mu g/m^3$ in the $PM_{2.5}$ level, the relative risk increases by 4.9%, see Figure 4.6). The relative risks for Non-Hispanic White are greater than one except for those who have

Table 4.9: Relative Risks for COPD complete models (*Significant at the 0.05 level).

Description			PM2.5 (increment=6 $\mu g/m^3$) Relative Risk (95% Wald CL)	NO2 (increment=17 ppb) Relative Risk (95% Wald CL)
Hispanic	Medicaid	Adults	1.119(0.913,1.371)	0.955(0.647,1.41)
		Elderly	1.182(0.948,1.475)	0.909(0.594,1.392)
	Medicare	Adults	0.994(0.915,1.08)	0.99(0.833,1.177)
		Elderly	1.051(0.979,1.128)	0.942(0.819,1.084)
	Private	Adults	0.974(0.836,1.135)	0.988(0.731,1.336)
		Elderly	1.029(0.88,1.205)	0.94(0.689,1.284)
	Uninsured	Adults	1(0.792,1.262)	1.467(0.979,2.199)
		Elderly	1.057(0.83,1.345)	1.396(0.914,2.132)
Non Hispanic -White	Medicaid	Adults	1.068(0.862,1.322)	1.077(0.713,1.628)
		Elderly	1.128(0.896,1.421)	1.025(0.653,1.611)
	Medicare	Adults	0.949(0.86,1.047)	1.117(0.917,1.36)
		Elderly	1.003(0.917,1.096)	1.063(0.889,1.271)
	Private	Adults	0.929(0.794,1.088)	1.114(0.816,1.522)
		Elderly	0.982(0.835,1.155)	1.06(0.764,1.471)
	Uninsured	Adults	0.954(0.747,1.22)	1.654(1.079,2.537)*
		Elderly	1.008(0.782,1.3)	1.574(1.005,2.467)*
Non Hispanic -Other	Medicaid	Adults	1.056(0.836,1.335)	1.302(0.808,2.099)
		Elderly	1.116(0.867,1.438)	1.239(0.744,2.065)
	Medicare	Adults	0.939(0.801,1.099)	1.35(0.983,1.855)
		Elderly	0.992(0.848,1.16)	1.285(0.946,1.745)
	Private	Adults	0.92(0.745,1.135)	1.347(0.89,2.04)
		Elderly	0.972(0.782,1.208)	1.282(0.837,1.963)
	Uninsured	Adults	0.944(0.72,1.238)	2(1.226,3.263)*
		Elderly	0.998(0.753,1.322)	1.903(1.145,3.163)*

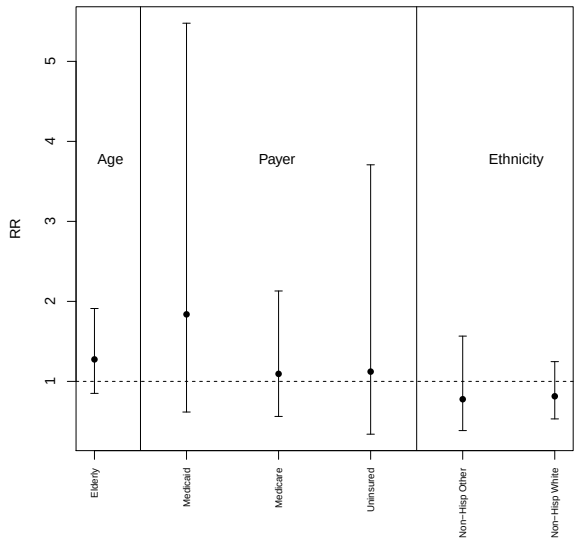


Figure 4.4: COPD complete model. Relative Risk and 95% Wald confidence limits. Each category is compared with its reference (see Table 4.8) at $PM_{2.5}$ value equal to 26.34 $\mu g/m^3$ (98th percentile).

Medicare. Finally for Non-Hispanic Others with Medicaid (RR 1.519) or Uninsured (RR 1.294) the relative risk of being hospitalized increases when the $PM_{2.5}$ increases by 6 $\mu g/m^3$. For NO_2 in the Non-Hispanic Other subgroup, Medicaid has a significant relative risk of 0.209(95% CL: 0.047,0.926). The relative risk and 95% confidence limits for Hispanic were plotted in Figure 4.7 . The relative risks for Non-Hispanic White and Non-Hispanic Other can be seen in Appendix A, Tables A.1-A.4.

Asthma Results

The lags for $PM_{2.5}$, NO_2, low wind speed and high wind speed are average of lag 0-1, average of lag 0-2, lag 3 and lag zero, respectively, see Chapter 3. From A.10, as with COPD, AT and low wind speed have a differential effect across subgroups that makes the reliability of the complete model suspect. In the complete model the parameters of the interaction between Adults and NO_2, as well as, the interaction between Elderly and NO_2 are negative and significant and NO_2 by itself is positive and significant, see Table A.11.

Table 4.10: Subgroup exploratory analyses for COPD.*Significant at 0.05 level.

Subgroup: Hispanic		
Description	Relative Risk (95% Wald CL)	
	PM2.5 (increment=6 $\mu g/m^3$)	NO2 (increment=17ppb)
Medicaid	0.934(0.714,1.222)	1.103(0.697,1.745)
Medicare	1.049(0.985,1.118)	0.935(0.821,1.065)
Private	0.937(0.773,1.136)	1.038(0.718,1.501)
Uninsured	0.915(0.699,1.198)	1.57(0.976,2.524)
Subgroup: Non-Hispanic White		
Description	Relative Risk (95% Wald CL)	
	PM2.5 (increment=6 $\mu g/m^3$)	NO2 (increment=17ppb)
Medicaid	1.288(0.839,1.976)	1.279(0.552,2.96)
Medicare	0.958(0.878,1.045)	1.113(0.934,1.327)
Private	1.061(0.856,1.315)	0.915(0.579,1.445)
Uninsured	1.343(0.741,2.434)	1.091(0.46,2.588)
Subgroup: Non-Hispanic Other		
Description	Relative Risk (95% Wald CL)	
	PM2.5 (increment=6 $\mu g/m^3$)	NO2 (increment=17ppb)
Medicaid	1.519(0.93,2.48)	0.209(0.047,0.926)*
Medicare	0.927(0.786,1.094)	1.365(0.986,1.89)
Private	0.557(0.211,1.468)	2.749(0.606,12.468)
Uninsured	1.294(0.617,2.715)	2.429(0.636,9.278)

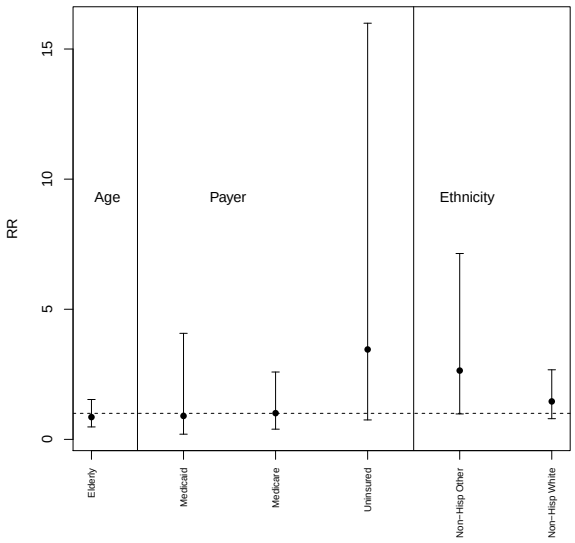


Figure 4.5: COPD complete model. Relative Risk and 95% Wald confidence limits. Each category is compared with its reference (see Table 4.8) at NO_2 value equal to 53.34 ppb (98th percentile).

For $PM_{2.5}$ the interaction with Medicare was positive and significant. Relative risk for the complete model are in Tables 4.11, 4.12 and 4.13; Children have the highest relative risks and Elderly the lowest relative risks for all categories in both $PM_{2.5}$ and NO_2 models.

Relative risks for complete model with NO_2

The relative risks for both Female and Male Children with Medicare are significantly higher than one for the three ethnicity categories when NO_2 increases by 17 pbb; Hispanic and Non-Hispanic White Female Children with Medicaid and Private Insurance have significantly higher risk of being hospitalized when the NO_2 increases by 17 pbb. For Hispanic Male Children and Non-Hispanic White Male or Female Children the risk is significantly higher than one if they have Private Insurance. In particular, Female Hispanic Children with Medicare are 1.554(95% CL 1.197-2.017) times more likely to be hospitalized due to asthma when NO_2 increases by 17 pbb. The relative risks for Young Adults with Medicare are significantly higher than one in all the categories, except for Non-Hispanic Other Male,

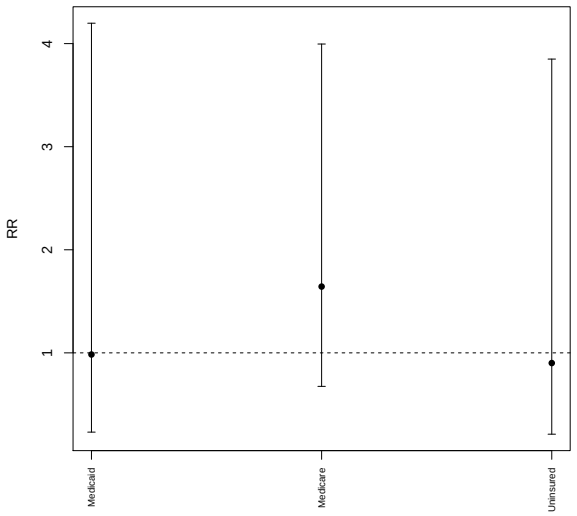


Figure 4.6: COPD Hispanic subgroup. Relative Risk and 95% Wald confidence limits. Each category is compared with its reference (see Table 4.8) at $PM_{2.5}$ value equal to 26.34 $\mu g/m^3$ (98th percentile).

see Tables 4.11, 4.12 and 4.13. The relative risk and 95% confidence limits are plotted in Figure (4.8) when the NO_2 level is equal to 55.41 ppb (98% percentile of the average of lag 0 to 2 for NO_2 values for cases and controls during the period $2005-2010$), we can see that the relative risk of Adults versus Children and Elderly versus Children are significantly less than one, people with Private Insurance are less likely to be hospitalized than people with Medicare, and Non-Hispanic Other and Non-Hispanic have higher probability of being hospitalized than Hispanic.

Relative Risks for complete model with $PM_{2.5}$

For $PM_{2.5}$ just the relative risks for Hispanic and Non-Hispanic Other with Medicare are significantly higher than one for Children and Young Adults when $PM_{2.5}$ increases by 6 $\mu g/m^3$. For Adult Hispanic Female with Medicare the relative risk is 1.155(95% CL 1.028-1.298). See, Tables 4.11, 4.12 and 4.13. The relative risk and 95% confidence limits are plotted in Figure (4.9) at $PM_{2.5}$ value of 24.77 $\mu g/m^3$ (98% percentile of the $PM_{2.5}$ average

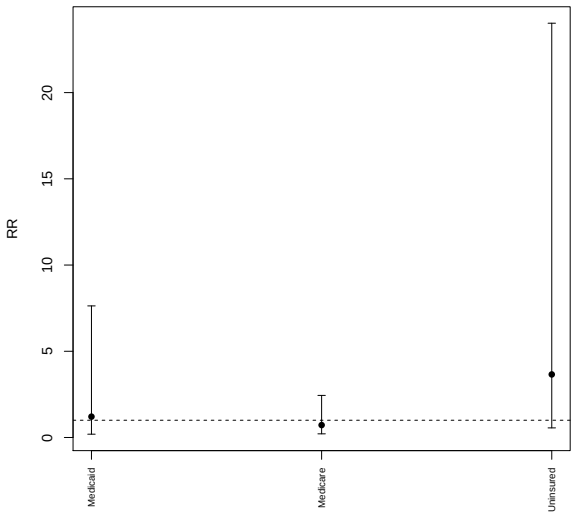


Figure 4.7: COPD Hispanic Subgroup. Relative Risk and 95% Wald confidence limits. Each category is compared with its reference (see Table 4.8) at NO_2 value equal to 53.34 ppb (98th percentile)

lag 0-1), we can see that the relative risk of people with Medicare versus people with Private Insurance is significantly greater than one.

Asthma Subgroup Analysis

When NO_2 increases by 17 ppb, the relative risk for Hispanic Children with Medicare or Private Insurance is significant greater than one, as in the complete model, Tables 4.11, 4.2.1. Unlike what we saw in the complete model with NO_2, there are no significant results for Children that are Non-Hispanic, but the relative risk for Non-Hispanic White Female Elderly on Medicare is significant. Non-Hispanic White Children with Medicare or Uninsured had the highest relative risks for a change in one IQR of NO_2 but not $PM_{2.5}$, although not significant at the 0.05 level; Non-Hispanic Other Young Adults on Medicare had the highest relative risks for a change in one IQR of either NO_2 or $PM_{2.5}$, (Tables 4.15, 4.16). Results for $PM_{2.5}$ are significant for Hispanic Children and Hispanic Young Adults with Medicare, see Table 4.2.1. From Figure 4.10 we cannot see any significant relative

Table 4.11: Relative Risks for Asthma complete models. Part I.* Significant at the 0.05 level of significance

Description				PM2.5 (increment=6 $\mu g/m^3$) Relative Risk (95% Wald CL)	NO2 (increment=17 ppb) Relative Risk (95% Wald CL)
Hispanics	Female	Medicaid	Children	1.065(0.957,1.186)	1.19(1.016,1.393)*
			Young Adults	1.063(0.918,1.232)	1.108(0.891,1.377)
			Adults	0.956(0.817,1.117)	0.905(0.722,1.136)
			Elderly	0.901(0.732,1.11)	0.871(0.644,1.177)
		Medicare	Children	1.287(1.071,1.547)*	1.554(1.197,2.017)*
			Young Adults	1.285(1.07,1.543)*	1.447(1.11,1.887)*
			Adults	1.155(1.028,1.298)*	1.182(0.998,1.401)
			Elderly	1.089(0.961,1.235)	1.138(0.943,1.373)
		Military	Children	1.202(0.868,1.663)	1.246(0.775,2.002)
			Young Adults	1.199(0.851,1.689)	1.16(0.701,1.92)
			Adults	1.078(0.759,1.53)	0.948(0.567,1.585)
			Elderly	1.017(0.698,1.481)	0.912(0.526,1.581)
		Private	Children	1.054(0.935,1.189)	1.256(1.048,1.506)*
			Young Adults	1.052(0.918,1.206)	1.169(0.952,1.436)
			Adults	0.946(0.813,1.101)	0.956(0.766,1.192)
			Elderly	0.892(0.728,1.093)	0.919(0.684,1.235)
		Uninsured	Children	1.061(0.882,1.276)	1.042(0.793,1.369)
			Young Adults	1.058(0.898,1.248)	0.97(0.756,1.246)
			Adults	0.951(0.792,1.143)	0.793(0.602,1.044)
			Elderly	0.897(0.717,1.124)	0.763(0.546,1.066)
	Male	Medicaid	Children	1.051(0.958,1.152)	1.134(0.989,1.3)
			Young Adults	1.049(0.891,1.234)	1.056(0.83,1.342)
			Adults	0.943(0.794,1.119)	0.863(0.672,1.108)
			Elderly	0.889(0.714,1.107)	0.83(0.605,1.14)
		Medicare	Children	1.27(1.063,1.516)	1.481(1.153,1.903)*
			Young Adults	1.267(1.04,1.544)	1.379(1.036,1.835)*
			Adults	1.139(0.992,1.309)	1.127(0.921,1.379)
			Elderly	1.075(0.93,1.241)	1.084(0.875,1.343)
		Military	Children	1.185(0.861,1.632)	1.187(0.741,1.903)
			Young Adults	1.183(0.834,1.677)	1.105(0.659,1.855)
			Adults	1.063(0.744,1.52)	0.903(0.533,1.532)
			Elderly	1.003(0.685,1.469)	0.869(0.496,1.525)
		Private	Children	1.04(0.934,1.158)	1.197(1.016,1.41)*
			Young Adults	1.038(0.89,1.21)	1.114(0.885,1.402)
			Adults	0.933(0.789,1.103)	0.911(0.712,1.164)
			Elderly	0.88(0.71,1.09)	0.876(0.642,1.196)
		Uninsured	Children	1.046(0.883,1.239)	0.993(0.771,1.279)
			Young Adults	1.044(0.88,1.239)	0.925(0.712,1.201)
			Adults	0.938(0.776,1.135)	0.756(0.567,1.007)
			Elderly	0.885(0.704,1.114)	0.727(0.516,1.024)

Table 4.12: Relative Risks for Asthma complete models. Part II. * Significant at the 0.05 level of significance

Description				PM2.5 (increment=6 $\mu g/m^3$) Relative Risk (95% Wald CL)	NO2 (increment=17 ppb) Relative Risk (95% Wald CL)
Non-Hisp White	Female	Medicaid	Children	1.028(0.863,1.224)	1.288(1,1.659)*
			Young Adults	1.026(0.843,1.248)	1.199(0.903,1.593)
			Adults	0.922(0.759,1.121)	0.98(0.739,1.3)
			Elderly	0.87(0.687,1.102)	0.943(0.672,1.324)
		Medicare	Children	1.242(0.991,1.558)	1.682(1.217,2.326)*
			Young Adults	1.24(0.994,1.546)	1.566(1.14,2.153)*
			Adults	1.115(0.949,1.308)	1.28(1.013,1.618)*
			Elderly	1.051(0.893,1.238)	1.232(0.969,1.565)
		Military	Children	1.16(0.825,1.63)	1.349(0.821,2.215)
			Young Adults	1.157(0.812,1.65)	1.256(0.748,2.109)
			Adults	1.04(0.727,1.488)	1.026(0.608,1.733)
			Elderly	0.981(0.67,1.437)	0.988(0.567,1.721)
		Private	Children	1.017(0.855,1.211)	1.36(1.052,1.758)*
			Young Adults	1.015(0.848,1.215)	1.266(0.973,1.647)
			Adults	0.913(0.76,1.096)	1.035(0.792,1.351)
			Elderly	0.861(0.687,1.078)	0.995(0.72,1.377)
		Uninsured	Children	1.023(0.811,1.291)	1.128(0.804,1.583)
			Young Adults	1.021(0.826,1.262)	1.05(0.77,1.434)
			Adults	0.918(0.737,1.144)	0.859(0.621,1.187)
			Elderly	0.866(0.672,1.116)	0.826(0.57,1.197)
	Male	Medicaid	1Children	1.014(0.862,1.192)	1.228(0.968,1.557)
			Young Adults	1.012(0.824,1.242)	1.143(0.848,1.541)
			Adults	0.91(0.741,1.116)	0.934(0.693,1.26)
			Elderly	0.858(0.673,1.094)	0.899(0.633,1.276)
		Medicare	Children	1.225(0.985,1.524)	1.603(1.172,2.194)*
			Young Adults	1.223(0.971,1.54)	1.493(1.07,2.083)*
			Adults	1.099(0.924,1.308)	1.22(0.945,1.576)
			Elderly	1.037(0.871,1.235)	1.174(0.907,1.519)
		Military	Children	1.144(0.819,1.598)	1.285(0.786,2.103)
			Young Adults	1.141(0.797,1.635)	1.197(0.704,2.035)
			Adults	1.026(0.714,1.475)	0.978(0.572,1.673)
			Elderly	0.968(0.658,1.423)	0.941(0.534,1.657)
		Private	Children	1.003(0.853,1.18)	1.296(1.017,1.65)*
			Young Adults	1.001(0.828,1.211)	1.206(0.912,1.596)
			Adults	0.9(0.742,1.093)	0.986(0.741,1.311)
			Elderly	0.849(0.673,1.072)	0.949(0.677,1.329)
		Uninsured	Children	1.009(0.812,1.255)	1.075(0.78,1.482)
			Young Adults	1.007(0.813,1.249)	1.001(0.728,1.376)
			Adults	0.906(0.725,1.132)	0.818(0.587,1.141)
			Elderly	0.854(0.662,1.102)	0.787(0.541,1.146)

Table 4.13: Relative Risks for Asthma complete models. Part III. * Significant at the 0.05 level of significance

Description				PM2.5 (increment=6 $\mu g/m^3$) Relative Risk (95% Wald CL)	NO2 (increment=17 ppb) Relative Risk (95% Wald CL)
Non-Hisp Other	Female	Medicaid	Children	1.066(0.896,1.267)	1.2(0.933,1.544)
			Young Adults	1.064(0.877,1.291)	1.117(0.843,1.482)
			Adults	0.956(0.782,1.169)	0.913(0.682,1.223)
			Elderly	0.902(0.703,1.156)	0.879(0.614,1.258)
		Medicare	Children	1.288(1.026,1.617)*	1.568(1.132,2.171)*
			Young Adults	1.285(1.03,1.604)*	1.459(1.061,2.008)*
			Adults	1.156(0.974,1.371)	1.193(0.93,1.531)
			Elderly	1.09(0.907,1.31)	1.148(0.877,1.502)
		Military	Children	1.202(0.852,1.695)	1.257(0.761,2.075)
			Young Adults	1.2(0.84,1.714)	1.17(0.693,1.975)
			Adults	1.079(0.749,1.554)	0.956(0.56,1.633)
			Elderly	1.017(0.687,1.506)	0.92(0.519,1.631)
		Private	Children	1.055(0.879,1.265)	1.267(0.969,1.657)
			Young Adults	1.053(0.874,1.268)	1.179(0.897,1.551)
			Adults	0.946(0.776,1.154)	0.964(0.722,1.287)
			Elderly	0.892(0.698,1.141)	0.928(0.651,1.322)
		Uninsured	Children	1.061(0.841,1.338)	1.051(0.749,1.475)
			Young Adults	1.059(0.858,1.307)	0.979(0.718,1.335)
			Adults	0.952(0.759,1.194)	0.8(0.573,1.117)
			Elderly	0.898(0.688,1.171)	0.77(0.522,1.136)
	Male	Medicaid	1Children	1.051(0.891,1.24)	1.144(0.899,1.455)
			Young Adults	1.049(0.853,1.29)	1.065(0.788,1.438)
			Adults	0.943(0.761,1.168)	0.87(0.637,1.189)
			Elderly	0.89(0.687,1.152)	0.837(0.577,1.215)
		Medicare	Children	1.27(1.016,1.588)*	1.494(1.087,2.053)*
			Young Adults	1.268(1.003,1.602)*	1.391(0.993,1.948)
			Adults	1.14(0.944,1.376)	1.137(0.865,1.494)
			Elderly	1.075(0.882,1.311)	1.094(0.82,1.46)
		Military	Children	1.186(0.844,1.666)	1.198(0.727,1.974)
			Young Adults	1.183(0.822,1.703)	1.115(0.651,1.909)
			Adults	1.064(0.733,1.543)	0.911(0.526,1.579)
			Elderly	1.003(0.674,1.495)	0.877(0.489,1.573)
		Private	Children	1.04(0.873,1.239)	1.207(0.933,1.563)
			Young Adults	1.038(0.85,1.268)	1.124(0.838,1.507)
			Adults	0.933(0.755,1.154)	0.919(0.674,1.251)
			Elderly	0.88(0.682,1.136)	0.884(0.612,1.278)
		Uninsured	Children	1.047(0.839,1.305)	1.002(0.725,1.385)
			Young Adults	1.044(0.841,1.298)	0.933(0.677,1.285)
			Adults	0.939(0.744,1.185)	0.762(0.54,1.076)
			Elderly	0.886(0.676,1.16)	0.734(0.493,1.09)

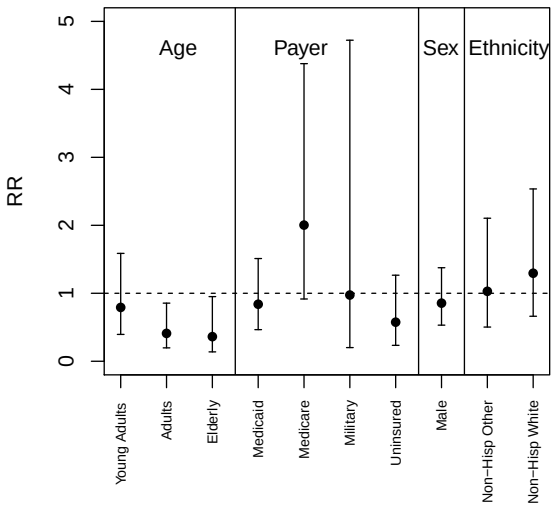


Figure 4.8: Asthma Relative Risk and 95% Wald confidence limits. Each category is compared with its reference (see Table 4.8) at NO_2 value equal to 55.41 ppb (98th percentile).

risk for NO_2, Hispanic Children are more likely to be hospitalized than Hispanic people in other age categories when NO_2 is equal to 55.41 ppb, we can see the same in Figure (4.11) when $PM_{2.5}$ is equal to 24.77 $\mu g/m^3$. See Appendix A (Tables A.5-A.8) for relative risk results for Non-Hispanic that were not plotted because the confidence intervals were very wide.

Other results

For COPD, see Table A.4, for Non-Hispanic Other, Medicaid versus Private has a RR of 0(CL:0,0.239) at the 98th percentile of NO_2. For Asthma, in Table A.8, for the Non-Hispanic Other subgroup, Uninsured versus Private had a RR of 0.044 (CL:0.003,0.722) at the 98th percentile of NO_2. For asthma Uninsured Non-Hispanic Others have lowest relative risks for one IQR change in NO_2 (Table 4.16), this was not the case for COPD (Table 4.10).

Table 4.14: Subgroup exploratory analyses for asthma (Hispanic).* Significant at the 0.05 level

Subgroup: Hispanic				
Description			Relative Risks (95% Wald CL)	
			PM2.5 (increment=6 $\mu g/m^3$)	NO2 (increment=17 ppb)
Female	Medicaid	Children	1.048(0.932,1.178)	1.154(0.97,1.372)
		Young Adults	1.022(0.866,1.205)	1.094(0.855,1.398)
		Adults	0.933(0.779,1.118)	0.929(0.717,1.204)
		Elderly	0.856(0.672,1.091)	0.86(0.607,1.218)
	Medicare	Children	1.281(1.032,1.59)*	1.437(1.063,1.942)*
		Young Adults	1.249(1.007,1.549)*	1.362(0.999,1.858)
		Adults	1.14(0.999,1.302)	1.157(0.956,1.401)
		Elderly	1.046(0.905,1.209)	1.071(0.861,1.332)
	Military	Children	1.088(0.712,1.663)	1.428(0.761,2.68)
		Young Adults	1.061(0.683,1.65)	1.354(0.705,2.6)
		Adults	0.969(0.619,1.517)	1.15(0.591,2.235)
		Elderly	0.889(0.553,1.431)	1.064(0.527,2.148)
	Private	Children	1.083(0.948,1.238)	1.226(1.003,1.5)*
		Young Adults	1.056(0.908,1.23)	1.163(0.92,1.469)
		Adults	0.965(0.81,1.148)	0.988(0.764,1.276)
		Elderly	0.885(0.699,1.12)	0.914(0.648,1.288)
	Uninsured	Children	1.087(0.888,1.331)	1.049(0.776,1.419)
		Young Adults	1.061(0.887,1.268)	0.995(0.753,1.315)
		Adults	0.968(0.788,1.19)	0.845(0.619,1.153)
		Elderly	0.889(0.688,1.148)	0.782(0.534,1.146)
Male	Medicaid	Children	1.058(0.959,1.167)	1.138(0.982,1.318)
		Young Adults	1.032(0.86,1.239)	1.079(0.823,1.413)
		Adults	0.942(0.771,1.151)	0.916(0.688,1.221)
		Elderly	0.865(0.67,1.115)	0.848(0.589,1.221)
	Medicare	Children	1.293(1.052,1.589)*	1.417(1.06,1.893)*
		Young Adults	1.261(1.004,1.585)*	1.343(0.964,1.872)
		Adults	1.152(0.984,1.348)	1.141(0.906,1.437)
		Elderly	1.057(0.897,1.246)	1.056(0.825,1.35)
	Military	Children	1.099(0.721,1.674)	1.408(0.752,2.635)
		Young Adults	1.072(0.683,1.681)	1.335(0.686,2.598)
		Adults	0.979(0.619,1.547)	1.134(0.575,2.237)
		Elderly	0.898(0.553,1.457)	1.049(0.514,2.142)
	Private	Children	1.094(0.972,1.231)	1.209(1.009,1.449)*
		Young Adults	1.067(0.899,1.266)	1.147(0.884,1.487)
		Adults	0.974(0.802,1.184)	0.974(0.732,1.296)
		Elderly	0.894(0.697,1.146)	0.901(0.629,1.292)
	Uninsured	Children	1.098(0.913,1.32)	1.035(0.783,1.368)
		Young Adults	1.071(0.887,1.292)	0.981(0.732,1.315)
		Adults	0.978(0.788,1.214)	0.833(0.601,1.157)
		Elderly	0.897(0.69,1.167)	0.771(0.521,1.141)

Table 4.15: Subgroup exploratory analyses for asthma (Non-Hispanic White) * Significant at the 0.05 level

Subgroup: Non-Hispanic White				
Description			Relative Risks (95% Wald CL)	
			PM2.5 (increment=6 $\mu g/m^3$)	NO2 (increment=17ppb)
Female	Medicaid	Children	1.227(0.785,1.919)	1.677(0.919,3.062)
		Young Adults	1.381(0.811,2.35)	1.349(0.647,2.815)
		Adults	1.303(0.798,2.128)	0.971(0.47,2.006)
		Elderly	1.465(0.814,2.637)	1.227(0.527,2.86)
	Medicare	Children	0.997(0.602,1.649)	2.067(0.997,4.287)
		Young Adults	1.122(0.669,1.879)	1.663(0.827,3.343)
		Adults	1.059(0.801,1.399)	1.197(0.791,1.813)
		Elderly	1.19(0.927,1.528)	1.512(1.037,2.206)*
	Military	Children	1.858(0.907,3.803)	1.481(0.555,3.948)
		Young Adults	2.09(0.952,4.589)	1.191(0.385,3.683)
		Adults	1.973(0.883,4.408)	0.858(0.274,2.683)
		Elderly	2.218(0.936,5.257)	1.084(0.322,3.651)
	Private	Children	0.86(0.591,1.25)	1.585(0.902,2.785)
		Young Adults	0.967(0.647,1.447)	1.275(0.755,2.155)
		Adults	0.913(0.636,1.31)	0.918(0.548,1.537)
		Elderly	1.026(0.634,1.661)	1.16(0.594,2.264)
	Uninsured	Children	0.895(0.448,1.789)	2.33(0.904,6.005)
		Young Adults	1.007(0.53,1.915)	1.874(0.805,4.362)
		Adults	0.951(0.5,1.806)	1.349(0.58,3.14)
		Elderly	1.069(0.523,2.185)	1.705(0.655,4.433)
Male	Medicaid	Children	1.024(0.69,1.519)	1.204(0.688,2.106)
		Young Adults	1.152(0.683,1.943)	0.969(0.452,2.076)
		Adults	1.087(0.684,1.73)	0.697(0.335,1.451)
		Elderly	1.222(0.689,2.169)	0.881(0.369,2.102)
	Medicare	Children	0.832(0.492,1.407)	1.484(0.715,3.081)
		Young Adults	0.936(0.53,1.653)	1.194(0.558,2.555)
		Adults	0.883(0.625,1.248)	0.859(0.53,1.395)
		Elderly	0.993(0.71,1.389)	1.086(0.67,1.761)
	Military	Children	1.55(0.785,3.06)	1.063(0.407,2.779)
		Young Adults	1.744(0.802,3.79)	0.855(0.27,2.705)
		Adults	1.646(0.751,3.606)	0.616(0.195,1.945)
		Elderly	1.85(0.791,4.328)	0.778(0.226,2.677)
	Private	Children	0.717(0.496,1.036)	1.138(0.653,1.983)
		Young Adults	0.807(0.52,1.251)	0.916(0.504,1.663)
		Adults	0.762(0.52,1.115)	0.659(0.375,1.158)
		Elderly	0.856(0.517,1.418)	0.833(0.403,1.722)
	Uninsured	Children	0.747(0.379,1.473)	1.673(0.676,4.136)
		Young Adults	0.84(0.436,1.619)	1.346(0.574,3.156)
		Adults	0.793(0.417,1.509)	0.969(0.42,2.233)
		Elderly	0.892(0.433,1.836)	1.224(0.467,3.204)

Table 4.16: Subgroup exploratory analyses for asthma (Non-Hispanic Other).* Significant at the 0.05 level of significance

Subgroup: Non-Hispanic Other				
Description			Relative Risks (95% Wald CL)	
			$PM_{2.5}$ (increment=6 $\mu g/m^3$)	NO2 (increment=17 ppb)
Female	Medicaid	Children	1.028(0.71,1.489)	1.062(0.632,1.786)
		Young Adults	1.118(0.758,1.65)	1.123(0.632,1.998)
		Adults	0.976(0.658,1.446)	0.91(0.495,1.672)
		Elderly	0.836(0.444,1.576)	0.663(0.259,1.694)
	Medicare	Children	1.387(0.809,2.379)	1.79(0.793,4.04)
		Young Adults	1.508(0.949,2.398)	1.893(0.945,3.792)
		Adults	1.316(0.958,1.808)	1.533(0.949,2.477)
		Elderly	1.128(0.727,1.75)	1.117(0.593,2.102)
	Military	Children	0.975(0.453,2.099)	0.876(0.293,2.621)
		Young Adults	1.06(0.457,2.46)	0.926(0.273,3.146)
		Adults	0.925(0.392,2.182)	0.75(0.215,2.615)
		Elderly	0.793(0.295,2.133)	0.546(0.13,2.295)
	Private	Children	0.967(0.657,1.425)	1.214(0.674,2.185)
		Young Adults	1.052(0.693,1.597)	1.283(0.702,2.347)
		Adults	0.918(0.576,1.462)	1.039(0.54,2)
		Elderly	0.787(0.401,1.545)	0.757(0.288,1.993)
	Uninsured	Children	0.884(0.458,1.707)	0.465(0.181,1.194)
		Young Adults	0.961(0.558,1.657)	0.492(0.232,1.043)
		Adults	0.838(0.47,1.495)	0.398(0.168,0.943)
		Elderly	0.719(0.337,1.535)	0.29(0.095,0.883)*
Male	Medicaid	Children	0.995(0.71,1.395)	1.169(0.733,1.864)
		Young Adults	1.082(0.67,1.748)	1.236(0.604,2.529)
		Adults	0.944(0.584,1.526)	1.001(0.477,2.101)
		Elderly	0.809(0.416,1.574)	0.729(0.268,1.989)
	Medicare	Children	1.342(0.797,2.259)	1.97(0.93,4.171)
		Young Adults	1.46(0.847,2.517)	2.083(0.95,4.565)
		Adults	1.274(0.834,1.945)	1.687(0.924,3.08)
		Elderly	1.092(0.671,1.775)	1.229(0.616,2.453)
	Military	Children	0.944(0.448,1.987)	0.964(0.322,2.885)
		Young Adults	1.026(0.425,2.476)	1.019(0.274,3.795)
		Adults	0.895(0.366,2.191)	0.826(0.217,3.147)
		Elderly	0.767(0.281,2.095)	0.601(0.135,2.683)
	Private	Children	0.936(0.665,1.318)	1.335(0.808,2.207)
		Young Adults	1.018(0.622,1.666)	1.412(0.694,2.871)
		Adults	0.888(0.522,1.511)	1.144(0.539,2.428)
		Elderly	0.761(0.379,1.529)	0.833(0.303,2.288)
	Uninsured	Children	0.855(0.483,1.515)	0.512(0.22,1.191)
		Young Adults	0.93(0.542,1.596)	0.541(0.246,1.192)
		Adults	0.811(0.458,1.437)	0.438(0.179,1.072)
		Elderly	0.696(0.335,1.443)	0.319(0.105,0.972)*

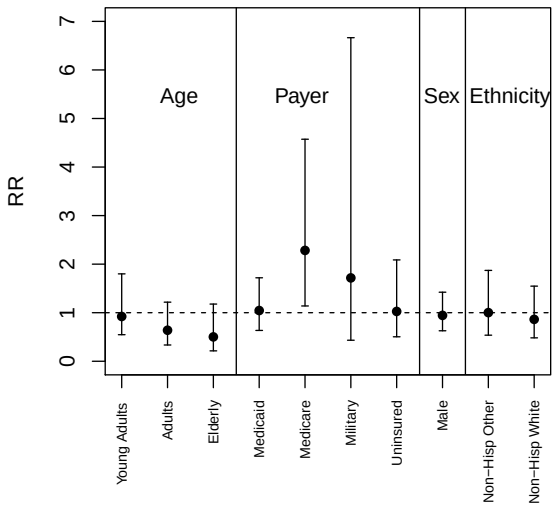


Figure 4.9: Asthma Relative Risk and 95% Wald confidence limits. Each category is compared with its reference (see Table 4.8) at $PM_{2.5}$ value equal to 24.77 $\mu g/m^3$ (98th percentile).

We described possible differential effects of pollution across subgroups and reported the parameter estimates (and standard errors), so that they can later be included as part of a comprehensive review as in *Provisional Assessment of Recent Studies on Health Effects of Particulate Matter Exposure (2006)* [32] and *Integrated Science Assessment for Oxides of Nitrogen Health Criteria (Final Report)* [31].

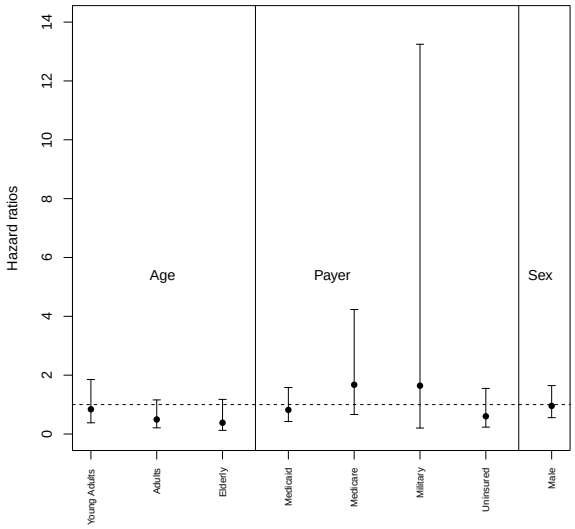


Figure 4.10: Hispanic Asthma Relative Risk and 95% Wald confidence limits. Each category is compared with its reference (see Table 4.8) at NO_2 value equal to 55.41*ppb* (98th percentile).

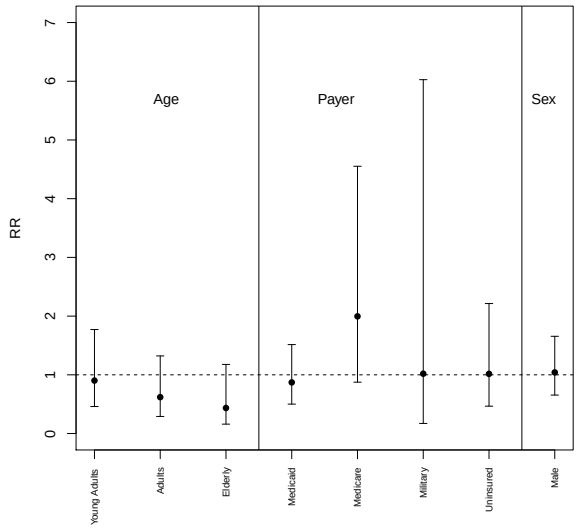


Figure 4.11: Hispanic Asthma Relative Risk and 95% Wald confidence limits. Each category is compared with its reference (see Table 4.8) at $PM_{2.5}$ value equal to 24.77 $\mu g/m^3$ (98th percentile).

Chapter 5

Conclusions

With the use of the historical functional linear model we avoid having to do a time consuming analysis for the choosing of the lags. Standard statistical software (R-package) was used for the calcultation by the modification of the COXPH function as in Bulathsinhala (2011) in order to allow us to use P-splines penalties instead of the ridge regression penalties. We guided the choose of the smoothing parameter by a modified ridge trace method, that worked for COPD and although for asthma the result was an oversmoothing curve, it was a good starting point for the search of the smoothing parameter. We have some advantages with the case cross-over design, since the subjects are their controls the confounding effects of seasonal variables are reduced, and we can work with social charateristics by including them as interaction terms. When we include interactions between a continuous variable and a categorical variable the parameter cannot be interpretated as the log of the odd ratios, however the PHREG function in SAS allows us to calculate the relative risk by specifing some values of the continuous (pollutant) variable.

We found that when $PM_{2.5}$ is equal to the 98% percentile of the daily values, Hispanics are more likely to be hospitalized due to asthma or COPD than Non-Hispanics and when NO_2 is equal to the 98% percentile for the daily values, in contrast to $PM_{2.5}$, Non-Hispanics are more likely to hospitalized due to asthma or COPD than Hispanics. For Hispanic Children with Medicare the probability of being hospitalized for asthma increases significantly when either $PM_{2.5}$ or NO_2 increases by one IQR. Additionally, Non-Hispanic Other who are uninsured have the lowest RR of asthma hospitalization, while those with Medicare have the highest when either NO_2 or $PM_{2.5}$ increase by one IQR. For COPD, Non-Hispanics have the highest RR for Medicaid and Uninsured for one IQR change in NO_2 or $PM_{2.5}$, but

for Hispanics, this is only true for changes in NO_2. We described possible differential effects of pollution across subgroups and reported the parameter estimates (and standard errors), so that they can later be included as part of a comprehensive review as in *Provisional Assessment of Recent Studies on Health Effects of Particulate Matter Exposure (2006)* [32] and *Integrated Science Assessment for Oxides of Nitrogen Health Criteria (Final Report)* [31].

One limitation in our study is that we had to assume that each hospitalization occurred at the end of the current day because the time of hospitalization is unavailable. Another limitation is that we assume that all the population in El Paso, TX is exposed to the same level to pollutants and weather conditions. Although we have six monitoring stations (CAMS) for pollutant in the El Paso, unfortunately we do not have enough data from most of them, and it is not feasible to match the patient home location (Zip Code) to the nearest CAMS.

References

[1] E. B. Andersen. Asymptotic Properties of Conditional Maximum-Likelihood Estimators. *Journal of the Royal Statistical Society. Series B*, Vol. 32(2):283-301, 1970.

[2] I. V. Basawa. Efficiency of Conditional Maximum Likelihood Estimators and Confidence Limits for Mixtures of Exponential Families. *Biometrika*, Vol. 68(2):515-523, 1981.

[3] V. Belledui, A. Faustini, M. Stafoggia, G. Cattani, A. Marconi, C. Perucci, F. Forastiere. Impact of Fine and Ultrafine Particles on Emergency Hospital Admissions for Cardiac and Respiratory Diseases. *Epidemiology*, Vol. 21(3):414-423, 2010.

[4] P. K. Bulathsinhala.Estimating the effect of dust and low wind events on hospitalizations for asthma while adjusting for hourly levels or air pollutants. Master's thesis, University of Texas at El Paso, 2011.

[5] T. A. Burra, Rahim Moineddin, Mohammad M. Agha, Richard H. Glazier, Social disadvantage, air pollution, and asthma physician visits in Toronto, Canada. *Environmental Research*, Vol. 109(5):567-574, 2009.

[6] P. H. C. Eilers, D. Marx. Flexible Smoothing with B-splines and Penalties, *Statistical Science*, Vol. 11(2):89-121,1996.

[7] R. L. Eubank.*Nonparametric Regression and Spline Smoothing*, Marcel Dekker,1988.

[8] S. E. Grineski, J. G. Staniswalis, P. Bulathsinhala, Y. Peng, and T. E. Gill. Hospital admissions for asthma and acute bronchitis in El Paso, Texas: Do age, sex, and insurance status modify the effects of dust and low wind events?. *Environmental Research*, Vol. 111(8):1148-1155, 2011.

[9] S. E. Grineski, J. G. Staniswalis, Y. Peng, and C. Atkinson-Palombo. Childrens asthma hospitalizations and relative risk due to nitrogen dioxide (no2): Efect modification by race, ethnicity and insurance status. *Environmental Research*, 110:178-188, 2010.

[10] D. W. Hosmer, S. Lemeshow. *Applied Logistic Regression*, John Wiley and Sons, Inc. U.S.A., 2nd edition, 2000.

[11] F. Huque, S. K. Katti, A Note on Maximum Conditional Likelihood Estimators. *Sankhyā: The indian Journal of Statistics , Series B*, Vol. 38(1):1-13, 1976.

[12] J. D. Kalbfleish, D. A. Sprott. Application of Likelihood Methods to Models Involving Large Numbers of Parameters. *Journal of the Royal Statistical Society. Series B*, Vol. 32(2):175-208, 1970.

[13] M. H. Kutner, C. J. Nachtsheim, J. Neter, W. Li. *Applied Linear Statistical Models*, McGraw-Hill/Irwin, 2005.

[14] K. Liang. On Information and Ancillarity in the Presence of a Nuisance Parameter. *Biometrika*, Vol. 70(3):607-612, 1983.

[15] K. Liang. The asymptotic efficiency of conditional likelihood methods. *Biometrika*, Vol. 71(2):305-313, 1984.

[16] F. W. Lipfert, Air pollution and poverty: does the sword cut both ways?. *Journal of Epidemiology and Community Health*, 58:2-3, 2004.

[17] M. Maclure, The Case-Crossover Design: A Method for Studying Transient Effects on the Risk of Acute Events. *American Journal of Epidemiology*, Vol. 133(2):144-153, 1991.

[18] National Heart, Lung and Blood Institute, `http://www.nhlbi.nih.gov/health/health-topics/topics/copd/`,Accessed:7/10/2013.

[19] B. Ostro, R. Broadwin, S. Green, W.-Y. Feng and M. Lipsett. Fine particulate air pollution and mortality in nine California counties: results from CALFINE. *Environ. Health Perspect*, 114:29-33, 2006.

[20] J. L. Peel, P. E. Tolbert, M. Klein, K. B. Metzger, W . D. Flanders, K. Todd, J. A. Mulholland, P. B. Ryan, H. Frumkin. Ambient air pollution and respiratory emergency department visits. *Epidemiology*, 16(2):164-74,2005.

[21] Y. Peng. A retrospective study of dust storms and respiratory hospitalizations in El Paso, Texas using a case- crossover study design. Master's thesis, University of Texas at El Paso, 2009.

[22] H. Qiu, I. T. S. Yu, L. W. Tian, X. R. Wang, L. A. Tse, W. Tam, T. W. Wong. Effects of Coarse Particulate Matter on Emergency Hospital Admissions for Respiratory Diseases: A Time-Series Analysis in Hong Kong.*Environmental Health Perspectives*, 120(4):572-76, 2012.

[23] N. Sartori, T. A. Severini. Conditional Likelihood Inference in Generalized Linear Mixed Models. *Statistica Sinica*, 14:349-360, 2004.

[24] SAS/STAT(R) 9.2 User's Guide, Second Edition, `http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_phreg_sect036.htm`, Accessed:03/22/2013

[25] C. O. Schmidt, T. Kohlmann. When to use the odds ratio or the relative risk?. *Int J Public Health* , 53:165-167,2008.

[26] D. A. Sprott, J. D. Kalbfleish. Examples of likelihoods and comparisons with point estimates and large sample approximations. Journal of the American Statistical Association, Vol. 64(326):468-484, 1969.

[27] J. G. Staniswalis, N. J. Parks, J. O. Bader, Y. M. Maldonado. Temporal analysis of airborne particulate matter reveals a dose-rate effect on mortality in El

Paso: indications of differential toxicity for different particle mixtures, *J. Air Waste Manage. Assoc.*, 55: 893-902,2005

[28] J. G. Staniswalis, H. Yang, W. Li, and K. E. Kelly. Using a continuous time lag to determine the association between ambient pm2.5 hourly levels and daily mortality. *Journal of the Air and Waste Management Association*, 59:1173-1185, 2009.

[29] R. Steadman. The assessment of sultriness. Part II: effects of wind, extra radiation and barometric pressure on apparent temperature. *Journal of Applied Meteorology*, 18:874-885,1979.

[30] J. M. Tenías, F. Ballester, M. L. Rivera. Association between Hospital Emergency Visits for Asthma and Air Pollution in Valencia Spain, *Occupational and Environmental Medicine*, Vol. 55(8):541-547,1998.

[31] U.S. EPA, Integrated Science Assessment for Oxides of Nitrogen Health Criteria (Final Report), *U.S. Environmental Protection Agency*, Washington, DC, EPA/600/R-08/071, 2008.

[32] U.S. EPA, Provisional Assessment of Recent Studies on Health Effects of Particulate Matter Exposure (2006), *U.S. Environmental Protection Agency*, Washington, DC, EPA/600/R-06/063, 2006.

[33] United States Enviromental Protection Agency, `http://www.epa.gov/airquality/particlepollution/`,Accessed:12/19/2012.

[34] United States Enviromental Protection Agency, `http://www.epa.gov/air/nitrogenoxides/`,Accessed:12/20/2012.

[35] United States Enviromental Protection Agency, `http://www.epa.gov/airquality/ozonepollution/`,Accessed:12/20/2012.

[36] United States Enviromental Protection Agency, `http://www.epa.gov/airquality/particlepollution/health.html`,Accessed:12/19/2012.

[37] S. V. Wang, B. A. Coull, J. Schwartz, M. A. Mittleman, G. A. Wellenius. Potential for Bias in Case-Crossover Studies With Shared Exposures Analyzed Using SAS. *American Journal of Epidemiology*, Vol. 174(1):118-124, 2011.

[38] H. Yang. A further study of the relationship between PM10 level and daily mortality in El Paso, TX using a historical functional linear model. Master's thesis, University of Texas at El Paso, 2005.

[39] J. Zhang, K. F. Yu. What's the Relative Risk?: A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes.*JAMA*, 280(19):1690-1691, 1998.

Appendix A

Tables

Table A.1: Relative Risk results for COPD when $PM_{2.5}$=26.34 $\mu g/m^3$. Non-Hispanic White subgroup.

Non-Hispanic White Subgroup: Relative Risks for Payer		
Description	**Point Estimate**	**95% Wald Confidence Limits**
Medicaid vs Private	2.343	(0.287,19.163)
Medicare vs Private	0.639	(0.232,1.763)
Uninsured vs Private	2.817	(0.176,45.216)

Table A.2: Relative Risk results for COPD when $PM_{2.5}$=26.34 $\mu g/m^3$. Non-Hispanic Other subgroup.

Non-Hispanic Other Subgroup: Relative Risks for Payer		
Description	**Point Estimate**	**95% Wald Confidence Limits**
Medicaid vs Private	81.771	(0.697,9588.42)
Medicare vs Private	9.371	(0.125,699.75)
Uninsured vs Private	40.491	(0.192,8523.80)

Table A.3: Relative Risk results for COPD when NO_2=53.34 ppb. Non-Hispanic White subgroup.

Non-Hispanic White Subgroup : Relative Risks for Payer		
Description	**Point Estimate**	**95% Wald Confidence Limits**
Medicaid vs Private	2.858	(0.143,57.07)
Medicare vs Private	1.85	(0.402,8.506)
Uninsured vs Private	1.737	(0.082,37.015)

Table A.4: Relative Risk results for COPD when NO_2=53.34 ppb. Non-Hispanic Other subgroup

Non-Hispanic Other Subgroup : Relative Risks for Payer		
Description	**Point Estimate**	**95% Wald Confidence Limits**
Medicaid vs Private	0	(0,0.239)
Medicare vs Private	0.111	(0.001,14.241)
Uninsured vs Private	0.678	(0.001,383.223)

Table A.5: Relative Risk results for asthma when $PM_{2.5}$=24.77 $\mu g/m^3$, Non-Hispanic White subgroup

Non-Hispanic White: Relative Risks		
Description	**Point Estimate**	**95% Wald Confidence Limits**
Payer Medicaid vs Private	4.351	(0.705,26.86)
Payer Medicare vs Private	1.843	(0.329,10.332)
Payer Military vs Private	24.085	(1.149,505.087)
Payer Uninsured vs Private	1.182	(0.085,16.443)
Young Adults vs Children	1.627	(0.23,11.521)
Adults vs Children	1.282	(0.215,7.649)
Elderly vs Children	2.077	(0.221,19.536)

Table A.6: Relative Risk results for asthma when $PM_{2.5}$=24.77 $\mu g/m^3$. Non-Hispanic Other subgroup

Non-Hispanic Other: Relative Risks		
Description	**Point Estimate**	**95% Wald Confidence Limits**
Male vs Female	0.874	(0.221,3.449)
Payer Medicaid vs Private	1.286	(0.254,6.51)
Payer Medicare vs Private	4.424	(0.531,36.893)
Payer Military vs Private	1.032	(0.038,28.157)
Payer Uninsured vs Private	0.688	(0.058,8.162)
Young Adults vs Children	1.414	(0.226,8.847)
Adults vs Children	0.805	(0.117,5.537)
Elderly vs Children	0.426	(0.028,6.564)

Table A.7: Relative Risk results for asthma when NO_2=55.41 ppb. Non-Hispanic White subgroup

Non-Hispanic White: Relative Risks		
Description	**Point Estimate**	**95% Wald Confidence Limits**
Male vs Female	0.34	(0.09,1.277)
Payer Medicaid vs Private	1.202	(0.146,9.909)
Payer Medicare vs Private	2.375	(0.358,15.748)
Payer Military vs Private	0.801	(0.025,25.366)
Payer Uninsured vs Private	3.508	(0.228,53.905)
Young Adults vs Children	0.492	(0.056,4.325)
Adults vs Children	0.169	(0.021,1.368)
Elderly vs Children	0.361	(0.029,4.578)

Table A.8: Relative Risk results for asthma when NO_2=55.41 ppb. Non-Hispanic other subgroup

Non-Hispanic Other: Relative Risks		
Description	**Point Estimate**	**95% Wald Confidence Limits**
Male vs Female	1.366	(0.29,6.445)
Payer Medicaid vs Private	0.648	(0.103,4.098)
Payer Medicare vs Private	3.549	(0.324,38.925)
Payer Military vs Private	0.346	(0.008,15.8)
Payer Uninsured vs Private	0.044	(0.003,0.722)
Young Adults vs Children	1.199	(0.141,10.178)
Adults vs Children	0.603	(0.063,5.791)
Elderly vs Children	0.215	(0.009,5.143)

Table A.9: Results of the parameter estimates for the weather parameters included on the single pollutant models for COPD.*Significant at 0.05 level

		Complete Model		Hispanics Subgroup		Non-Hispanics White Subgroup		Non-Hispanics Other	
Pollutant Model	**Covariates**	**Parameter Estimate (std error)**	**Relative Risk (95% Wald CL)**	**Parameter Estimate (std error)**	**Relative Risk (95% Wald CL)**	**Parameter Estimate (std error)**	**Relative Risk (95% Wald CL)**	**Parameter Estimate (std error)**	**Relative Risk (95% Wald CL)**
$PM_{2.5}$	lag1at_0	0.262(0.227)	1.299(0.832,2.028)	0.427(0.297)	1.532(0.857,2.741)	-0.029(0.397)	0.971(0.446,2.115)	0.006(0.755)	1.006(0.229,4.417)
NO_2		0.306(0.232)	1.358(0.862,2.138)	0.411(0.302)	1.508(0.834,2.728)	0.062(0.407)	1.064(0.479,2.365)	0.256(0.759)	1.292(0.292,5.722)
$PM_{2.5}$	lag1at_1	0.212(0.188)	1.236(0.854,1.788)	0.424(0.244)	1.528(0.947,2.465)	-0.032(0.332)	0.969(0.505,1.856)	-0.446(0.637)	0.64(0.184,2.233)
NO_2		0.241(0.191)	1.273(0.876,1.85)	0.407(0.247)	1.502(0.926,2.438)	0.030(0.338)	1.031(0.531,2)	-.230(0.638)	0.795(0.228,2.774)
$PM_{2.5}$	lag1at_2	0.067(0.169)	1.07(0.768,1.49)	0.185(0.218)	1.203(0.784,1.846)	0.027(0.298)	1.028(0.573,1.842)	-0.771(0.594)	0.463(0.144,1.482)
NO_2		0.090(0.170)	1.094(0.783,1.528)	0.171(0.220)	1.187(0.771,1.827)	0.067(0.301)	1.07(0.594,1.928)	-0.574(0.595)	0.563(0.175,1.808)
$PM_{2.5}$	lag1at_3	0	.	0	.	0	.	0	.
NO_2		0	.	0	.	0	.	0	.
$PM_{2.5}$	Low WS	0.080(0.061)	1.083(0.962,1.22)	0.036(0.079)	1.037(0.887,1.211)	0.219(0.105)*	1.245(1.014,1.528)	-0.170(0.228)	0.844(0.54,1.318)
NO_2		0.079(0.060)	1.082(0.961,1.218)	0.025(0.079)	1.026(0.879,1.197)	0.230(0.104)*	1.278(1.026,1.543)	-0.170(0.229)	0.843(0.539,1.32)
$PM_{2.5}$	High WS	0.042(0.063)	1.043(0.923,1.179)	0.080(0.082)	1.083(0.923,1.271)	0.007(0.111)	1.007(0.81,1.252)	-0.025(0.204)	0.975(0.653,1.456)
NO_2		0.042(0.063)	1.043(0.922,1.18)	0.083(0.082)	1.086(0.925,1.275)	-.001(0.111)	0.999(0.804,1.242)	-0.036(0.205)	0.964(0.645,1.441)

Table A.10: Results of the parameter estimates for the weather parameters included in the single pollutant models for Asthma. *Significant at 0.05 level

		Complete Model		Hispanics Subgroup		Non-Hispanics White Subgroup		Non-Hispanics Other	
Pollutant Model	**Covariates**	**Parameter Estimate (std error)**	**Relative Risk (95% Wald CL)**	**Parameter Estimate (std error)**	**Relative Risk (95% Wald CL)**	**Parameter Estimate (std error)**	**Relative Risk (95% Wald CL)**	**Parameter Estimate (std error)**	**Relative Risk (95% Wald CL)**
PM2.5	lag1at_0	-0.012(0.207)	0.988(0.658,1.483)	0.135(0.237)	1.145(0.72,1.821)	-0.499(0.567)	0.607(0.2,1.846)	-0.434(0.692)	0.648(0.167,2.516)
NO2		-0.001(0.208)	0.999(0.665,1.501)	0.151(0.237)	1.163(0.731,1.851)	-0.520(0.5676)	0.595(0.196,1.809)	-0.458(0.693)	0.633(0.163,2.462)
PM2.5	lag1at_1	0.348(0.173)*	1.417(1.009,1.99)	0.545(0.197)	1.725(1.172,2.54)	-0.471(0.463)	0.624(0.252,1.547)	-0.034(0.605)	0.967(0.296,3.163)
NO2		0.315(0.173)	1.371(0.976,1.925)	0.511(0.197)*	1.668(1.133,2.455)	-0.501(0.461)	0.606(0.246,1.494)	-0.104(0.607)	0.901(0.274,2.96)
PM2.5	lag1at_2	0.062(0.162)	1.064(0.774,1.462)	0.241(0.185)	1.274(0.886,1.832)	-0.905(0.427)*	0.405(0.175,0.935)	0.026(0.560)	1.026(0.343,3.073)
NO2		0.06044(0.162)	1.062(0.773,1.46)	0.239(0.185)	1.27(0.883,1.827)	-0.923(0.426)*	0.397(0.172,0.914)	0.011(0.564)	1.012(0.335,3.053)
PM2.5	lag1at_3	0	.	0	.	0	.	0	.
NO2		0	.	0	.	0	.	0	.
PM2.5	Low WS lag 3	0.00455(0.050)	1.005(0.911,1.107)	0.001(0.056)	1.001(0.897,1.118)	0.113(0.145)	1.119(0.842,1.488)	-0.111(0.160)	0.895(0.655,1.224)
NO2		-0.008(0.050)	0.992(0.9,1.094)	-0.010(0.056)	0.99(0.887,1.106)	0.101(0.145)	1.106(0.833,1.47)	-0.109(0.160)	0.896(0.655,1.226)
PM2.5	High WS	0.074(0.054)	1.077(0.968,1.198)	0.056(0.062)	1.057(0.937,1.194)	0.208(0.156)	1.231(0.908,1.67)	0.050(0.172)	1.052(0.751,1.473)
NO2		0.117(0.056)	1.124(1.008,1.254)	0.096(0.064)	1.101(0.972,1.247)	0.268(0.161)	1.307(0.954,1.792)	0.062(0.175)	1.064(0.755,1.501)

Table A.11: Results of the parameter estimates for the interaction and pollutant parameters included in the Asthma and COPD models (Standard Error).
* Significant at 0.05 level. **Significant at 0.10 level (Wald Test).

Asthma						
	avno2lag02*et	Non-Hisp Other	0.001(0.007)	NA	NA	NA
	avno2lag02*et	Non-Hisp White	0.005(0.006)	NA	NA	NA
	avno2lag02*SEX_CODE	Male	-0.003(0.004)	-0.001(0.005)	-0.019(0.012)	0.006(0.014)
	avno2lag0*Payer	Medicaid	-0.003(0.005)	-0.004(0.006)	0.003(0.019)	-0.008(0.017)
	avno2lag0*Payer	Medicare	0.013(0.007)**	0.009(0.009)	0.016(0.017)	0.023(0.022)
	avno2lag0*Payer	Military	0(0.015)	0.009(0.019)	-0.004(0.032)	-0.019(0.035)
	avno2lag0*Payer	Uninsured	-0.011(0.008)	-0.009(0.009)	0.023(0.025)	-0.056(0.026)*
NO2	avno2lag02*ag	Adults	-0.016(0.007)*	-0.013(0.008)	-0.032(0.019)	-0.009(0.021)
	avno2lag02*ag	Elderly	-0.018(0.009)*	-0.017(0.010)	-0.018(0.023)	-0.028(0.029)
	avno2lag02*ag	Young Adults	-0.004(0.006)	-0.003(0.007)	-0.013(0.02)	0.003(0.02)
	avno2lag02	NO2 avg lag 0-2	0.013(0.005)*	0.012(0.006)	0.027(0.017)	0.011(0.018)
	avPMlag01*et	Non-Hisp Other	0(0.013)	NA	NA	NA
	avPMlag01*et	Non-Hisp White	-0.006(0.012)	NA	NA	NA
	avPMlag01*SEX_CODE	Male	-0.002(0.008)	0.002(0.010)	-0.03(0.025)	-0.005(0.028)
	avPMlag01*pay_status	Medicaid	0.002(0.01)	-0.006(0.011)	0.059(0.037)	0.01(0.033)
	avPMlag01*pay_status	Medicare	0.033(0.014)*	0.028(0.017)	0.025(0.036)	0.06(0.044)
	avPMlag01*pay_status	Military	0.022(0.028)	0.001(0.037)	0.128(0.063)*	0.001(0.068)
	avPMlag01*pay_status	Uninsured	0.001(0.015)	0.001(0.016)	0.007(0.054)	-0.015(0.051)
PM2.5	avPMlag01*ag	Adults	-0.018(0.013)	-0.019(0.016)	0.01(0.037)	-0.009(0.04)
	avPMlag01*ag	Elderly	-0.028(0.018)	-0.034(0.021)	0.03(0.046)	-0.034(0.056)
	avPMlag01*ag	Young Adults	0(0.012)	-0.004(0.014)	0.02(0.04)	0.014(0.038)
	avPMlag01	PM2.5 avg lag 0-1	0.009(0.01)	0.013(0.011)	-0.025(0.032)	-0.006(0.033)
COPD						
	lag36no2*et	Non Hispanics Other	0.018(0.01)**	NA	NA	NA
	lag36no2*et	Non Hispanics White	0.007(0.006)	NA	NA	NA
	lag36no2*ag	Elderly	-0.003(0.014)	NA	NA	NA
	lag36no2*pay_status	Medicaid	-0.002(0.009)	0.004(0.018)	0.02(0.029)	-0.152(0.064)*
NO2	lag36no2*pay_status	Medicare	0(0.015)	-0.006(0.012)	0.012(0.015)	-0.041(0.046)
	lag36no2*pay_status	Uninsured	0.023(0.006)	0.024(0.018)	0.01(0.029)	-0.007(0.061)
	lag36no2	Avg. Lag 3-6 No2	-0.001(0.009)	0.002(0.011)	-0.005(0.014)	0.059(0.045)
	lag6pm25*et	Non Hispanics Other	-0.01(0.014)	NA	NA	NA
	lag6pm25*et	Non Hispanics White	-0.008(0.008)	NA	NA	NA
	lag6pm25*ag	Elderly	0.009(0.021)	NA	NA	NA
	lag6pm25*pay_status	Medicaid	0.023(0.013)	-0.001(0.028)	0.032(0.041)	0.167(0.092)**
PM2.5	lag6pm25*pay_status	Medicare	0.003(0.023)	0.019(0.017)	-0.017(0.02)	0.085(0.083)
	lag6pm25*pay_status	Uninsured	0.004(0.008)	-0.004(0.028)	0.039(0.054)	0.141(0.104)
	lag6pm25	Lag 6 PM2.5	-0.004(0.013)	-0.011(0.016)	0.01(0.018)	-0.098(0.082)

Curriculum Vitae

Juana Maribel Herrera Hernandez was born on July 29, 1981 in Cd. Juarez Chihuahua Mexico. She entered Instituto Tecnologico de Cd. Juarez in the fall of 1999 and received her bachelor's degree in Industrial Engineering in the summer of 2004. In the fall of 2005, she entered Universidad Autonoma de Cd. Juarez and completed the Master of Engineering in Manufacturing Program. From 2005 to 2010 she worked as a Manufacturing Engineer in Mexico.

In the spring of 2011, she entered the Graduate School of The University of Texas at El Paso. While pursuing a master's degree in Statistics she worked as a Teaching and Research Assistant. Maribel wishes to apply the knowledge gained from her studies in the fields of Engineering and Statistics.

Permanent address: 4959 De los Dorados
Cd. Juarez, Chihuahua, Mexico 32000