UNIVERSITY OF CINCINNATI

Aug 13th, 2003

I, Wei Zhong,

hereby submit this as part of the requirements for the degree of:

Master of Science

in Department of Environmental Health

It is entitled Analysis of Short-term Influences of Ambient Pollen Grains and Fungal Spores on Asthma Hospital Visits in Cincinnati, OH

Approved by:

Rakesh Shukla

Linda Levin

Gina Ogilvie

Analysis of Short-term Influences of Ambient Pollen Grains and Fungal Spores on Asthma Hospital Visits in Cincinnati, OH

A thesis submitted to the

Division of Research and Advanced Studies

of the University of Cincinnati

in partial fulfillment of the
Requirements for the degree of

MASTER OF SCIENCE

in the Department of Environmental Health

of the College of Medicine

2003

by

Wei Zhong
B. S., Shanghai Medical University, 1999

Committee Chair: Dr. Rakesh Shukla

Abstract

Generalized linear models and nonparametric generalized additive models were employed to investigate the influences of pollen grains and fungal spores on asthma hospital visits in Cincinnati.

In the generalized linear model, indicator variables for each month and restricted cubic splines of time term were used to filter out the seasonal variation and time trends in asthma visits. Temperature and humidity were treated categorically to take out the asthma dependence on weather. Autoregressive Poisson models were estimated if serial correlation detected in the residuals.

The Generalized additive model is a more flexible approach when the nature of the asthma dependence on covariates is unknown. Counts of daily asthma hospital visits were fitted to nonparametric smoothing splines of time, temperature, and humidity and a linear term of aeroallergens.

Results from both models indicated that significant lag effects of total pollen, Oak/Cedar pollen and *Cladosporium* were 2, 3 and 4 days, respectively. Because of the superiority of the generalized additive model, 5 day lag effects were also found significant for *Alternaria* and Ragweed pollen .Plots of relative risk were generated and assessed for significant aeroallergens.

Acknowledgments

I would like to express my gratitude to Dr. Rakesh Shukla and Dr. Linda Levin for their inspiration, professional guidance and support for my thesis. I am indebted to their constant encouragement and mentoring throughout my graduate study.

I would also like to thank Dr. Tiina Reponen and Dr. Gurjit K. Khurana Hershey for serving on my committee and allowing me to use the data they have collected. Dr. Reponen provided many insightful suggestions and made contributions to this research as a bioaerosol specialist. Dr. Hershey helped in interpreting the results from medical perspective with her expertise in childhood asthma. Special thanks are extended to Dr. Atin Adhikari at the Center for Health-Related Aerosol Studies at Environmental Health Department. Dr. Adhikari gave me constant professional support whenever I went to him for advice. His sincere help is much appreciated.

Finally I would like to convey my deepest gratitude to my parents. I'm so grateful for their unconditional love and long-time support, without them, I would not have achieved so far in my path to knowledge and wisdom.

Table of Contents

Appendix

List of Tables

Table Page

List of Figures

Chapter1. Introduction

1.1 Asthma Overview

Asthma is defined as a chronic inflammatory pulmonary disorder that is characterized by reversible obstruction of the airways. Since the 1980s, there has been a worldwide increase in the prevalence of asthma in both children and adults. This escalating prevalence has led to significant increases in morbidity and mortality due to the disease. Approximately 15 million people in the United States have asthma, 5000 of whom die annually from the disease [1]. Asthma is now one of the most commonly reported diseases of childhood [2]. More children suffer from asthma than any other chronic disease, and the proportion of children reported to have asthma, to be hospitalized for asthma, or to die of asthma has increased substantially during the last two decades [3]. An estimated 5.8% of children under age 18, or 3.7 million children, have asthma [4].

The underlying cause of the increasing prevalence of asthma is unknown. However, the airway inflammation that is noted in asthma is due to an immune-mediated process in which inflammatory cells and inflammatory mediators enter airway tissues to cause disease. Many cell-mediated immunologic factors participate in the inflammatory process of asthma. The most important inflammatory cells involved are eosinophils, mast cells, and T lymphocytes.

1.2 The Effect of Outdoor Air Pollutants and Aeroallergens on Asthma

Long-time studies have described asthma as a result of the interaction between genetic background and the environment. However, the increased prevalence has occurred in too short a time to attribute it to genetic changes, and the identification of environmental factors is important to develop strategies for prevention.

Theoretically, there is a variety of ways in which environmental air pollutants could affect the prevalence and/or severity of asthma [5, 6]. They could act to induce, or augment directly, airway inflammation and hyperresponsiveness and some of them may have the ability to augment, or in other ways modify, immune responses to inhaled antigens such as to facilitate allergic sensitization or to enhance the severity of allergic reactions. Many studies have examined the relationship between air pollutants and asthma. Some of these have found that inhalation of air pollutants such as ozone, nitrogen oxides, carbon monoxide, sulfur dioxide and particulate matter, either individually or in combination, can enhance the airway response to inhaled allergens in atopic subjects inducing asthma exacerbations. Schwartz [7] found association between inhalable particles and ozone and hospital admissions for respiratory illness in the elderly from January 1, 1986 to December 31,1989 in Detroit. Slater et al. [8] confirmed that exposure to respirable particulate matter (PM_{10}) was a risk factor for the exacerbation of asthma after studying the daily records from September 1, 1989 to September 30, 1990 of asthma emergency room visits from eight hospitals in Seattle. In a study carried out at three cities in Ohio - Cincinnati, Cleveland, and Columbus [9], it was found that sulfur dioxide, ozone, and nitrogen dioxide were significantly associated with emergency department visits for asthma from 1991 to 1996. Ganor et al. [10] found that high nitrogen oxides and sulfur dioxide were positively correlated with emergency room visits of asthmatic children from January 1 to December 31, 1993 in Tel Aviv, Israel.

Besides air pollutants, the atmospheric environment also contains a multitude of aeroallergens such as pollen and fungal spores. Pollens are produced by plants in order to reproduce themselves and have a seasonal existence in the air. In contrary, large numbers

of airborne fungal spores are usually present in outdoor air throughout the year except winter and frequently exceeding pollen concentration by 100- to 1000-fold, depending on environmental factors such as humidity, temperature, and wind [11]. It has been known for some time that pollen and fungal spores, or at least components of them, can trigger symptoms of allergic respiratory diseases such as asthma and hay fever. In the clinical setting, skin-prick test and intradermal test using extracts of pollen and/or fungal spore allergens can determine atopic characteristics. Blood tests such as the raido-allergosorbent test (RAST) and multiple allergosorbent chemilluminescent assay (MAST_CLA) can also indicate allergy to pollen and fungal spore allergens [12,13]. The most common types of allergenic fungal spores are *Alternaria, Cladosporium* and the *Aspergillus* species, while the types of pollen that most commonly cause allergic reactions are produced by plain-looking plants (trees, grasses, and weeds) that do not have showy flowers.

Most of the studies done so far to examine the relationship of pollen and fungal spore levels with asthma are clinical studies. Muno et al [14] found that a group of children with rye-grass pollen lung sensitivity has a significant exacerbation in their asthma during the pollen season in Melbourne, Australia. Brown and Jackson [15] found associations of grass pollen counts and *Alternaria*, *Cladosporium* spores counts with asthma symptom severity among patients attending a Derby clinic. Reid et al. [16] found strong correlation between grass pollen counts and asthma emergency room visits and admissions in inland northern California. In a study of nine subjects with pollen-sensitized seasonal asthma in Quebec, Canada, Boulet et al. [17] found an increase in asthma symptom scores and reduced peak expiratory flow rates during the pollen season. However, in spite of the

striking parallelism of seasonal variations in asthma hospitalizations to seasonal trends in environmental pollen and fungal levels, quantitative epidemiological studies comparing measures of asthma symptoms (e.g. hospital admissions and attendances) with pollen and fungal spore concentrations are far less frequent than those that have been carried out in inorganic air pollutants and asthma research. Among those who have investigated the association of asthma with pollen and fungal spore concentrations quantitatively, Rosas et al. [18] found that relationship between fungal spore concentrations and asthma emergency admissions lagged 5 days in the year of 1991 in a Mexico City hospital. Lierl et al. [19] found significant association of the daily pollen concentration, but insignificant association of daily fungal spore concentration, with the number of asthma visits lagging 3 days in Cincinnati Children's Hospital Medical Center for the months of April through October in 1996 and 1997. The reason to model the lag day effects is that poor air quality may not always result in an immediate asthma visit, so air quality measures from 0 to 5 days before the asthma visit date needs to be considered to account for any delayed visit effects. All the above mentioned studies used Rotorod Sampler which is not the optimal method for measuring airborne pollen and fungal spores, as the smaller particles tend to flow around the sampling rod rather than impacting on it and thus can't be collected. Meanwhile, the unavailable information of individual types of pollen and fungal spores made it impossible to seek the relationship of asthma exacerbation with specific pollen and fungal spores.

1.3 Statistical Approaches in Research on the Relationship between Asthma and Outdoor Air Pollutants and Aeroallergens

Most epidemiology studies have used daily number of hospital asthma visits and daily concentration of air pollutants and aeroallergens to determine the relationship of asthma and particular environmental pollutants. The advantage of using time series of events is that they are unlikely to be confounded by personal risk factors such as smoking, blood pressure, and socioeconomic factors, since these factors do not vary from day to day with air pollutants.

1.3.1. *Generalized linear model*

The fraction of the population visiting a hospital for asthma on a given day is quite small, and the daily admission is usually expressed as a count number, which can only take on values limited to the nonnegative integers. This suggests that a Poisson stationary process is the underlying mechanism being modeled. Ordinary least-squares regression can not be used because it assumes that conditional on the explanatory factors, the outcome is normally distributed.

Poisson regression analysis assumes that conditional on the explanatory variables, the counts of events follow Poisson distribution. It belongs to the class of generalized linear models, which is an extension of traditional linear models that allows the mean of a population to depend on a linear predictor through a nonlinear link function and allow the response probability distribution to be any member of an exponential family of distributions. A generalized linear model consists of the following components:

- The linear component is defined just as it is for traditional linear models:

$$\eta = X\beta \qquad (1)$$

- A monotonic differentiable link function g describes how the expected value μ of Y is related to the linear predictor η:

$$g(\mu)=\eta=X\beta \qquad (2)$$

- The response variables Y are independent and have a probability distribution from an exponential family.

The general form of the log likelihood of the exponential family, share a common form: $l(\theta,\phi;y)=\frac{y\theta-b(\theta)}{\phi}+c(y,\phi)$, where θ is the natural parameter and ϕ is the dispersion parameter. The natural parameter is often written θ(μ) because it is a function of the expected value, μ. It can be shown that $\mu=\frac{\partial b(\theta)}{\partial\theta}$ and $\text{var}(y)=\phi\frac{\partial^2 b(\theta)}{\partial\theta^2}$ [20]. The second derivative, $\frac{\partial^2 b(\theta)}{\partial\theta^2}$, is usually denoted $V(\mu)$ and is called the variance function. Thus, $\text{var}(y)=\phi V(\mu)$, illustrating the dependence of the variance on the scale parameter, through ϕ, and the mean, through $V(\mu)$. The general form of the log likelihood provides one common rationale for selecting a suitable link function for a given distribution. The observations, denoted by the random variable y, are linear in the natural parameter, θ. Therefore, it would make sense to fit a linear model to θ rather than directly to the expected value, μ. For the Poisson distribution, the probability distribution is $f(y)=\frac{\lambda^y e^{-\lambda}}{y!}$ with mean equal to λ, the log likelihood function is thus $l(\lambda;y)=y\log(\lambda)-\lambda-\log(y!)$, and rewriting it in the common form of the log likelihood of the exponential family, we could see that $\theta=\log(\lambda)=\log(E(Y))$, $\phi=1$, $\text{var}(y)=\phi\frac{\partial^2 b(\theta)}{\partial\theta^2}=\lambda$, thus it would be reasonable to assume:

$$\theta = \log(E(Y)) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \qquad (3)$$

where Y is the count of hospital visit on a given day, E(Y) is the expected value of Y on that day, $X_1 \cdots X_p$ are the predictors of specific air pollutant's daily concentration and $\beta_1 \cdots \beta_p$ are the regression coefficients for those predictors.

The maximum-likelihood estimates of the parameters β in the linear predictor η can be obtained by iterative weighted least squares. Fitted generalized linear models can be summarized through statistics such as parameter estimates, their standard errors. Influence can also be made about the parameters using confidence intervals and hypothesis tests. However, specific inference procedures are usually based on asymptotic considerations, since exact distribution theory is not available or is not practical for all generalized linear models.

The deviance is defined as $2[l(\theta(y); y) - l(\theta(X\hat{\beta}); y)]$ [20], where $l(\theta(y); y)$ is the log likelihood, with $\theta(y)$ equals to the value of θ determined from the data and $\theta(X\hat{\beta})$ determined from the estimate of β under the model. The deviance has an approximate χ^2 distribution with *N-P* degrees of freedom, where *N*=the number of observations and *P* is the rank of *X*. For distributions that do not depend on an unknown scale parameter, such as the Poisson, the deviance tests lack of fit of the model. A test of the hypothesis of the significance of each additional term fit can be carried out by a Type I analysis, which is sometimes called an analysis of deviance. A type I analysis consists of fitting a sequence of models, beginning with a simple model with only an intercept term, and continuing through a model of specified complexity, fitting one additional effect on each step. Likelihood ratio statistics, that is, twice the difference of the log likelihoods, are computed between successive models. If the dispersion parameter is held fixed for all

models, it is equivalent to computing differences of scaled deviances. The asymptotic distribution of the likelihood ratio statistics, under the hypothesis that the additional parameters included in the model are equal to 0, is a chi-square with degrees of freedom equal to the difference in the number of parameters estimated in the successive models.

1.3.2. *Generalized Additive Models*

For environmental problems, the functional form of the dependence of the outcome on covariates and risk factor is often unknown and hard to estimate. However, if one is to examine the association of an environmental risk factor with the outcome, and there is any possibility that it is correlated with the covariates, proper control for the covariate is necessary if we are to avoid hopeless confounding. Therefore, in the absence of any theory to guide us, a flexible approach to covariate control is appropriate, and generalized additive model [21] is a desirable alternative. It is an extension of generalized linear model, and instead of assuming expected value of the outcome on a sum of linear functions of predictors, the outcome is assumed to depend on a sum of smooth functions of the predictors [21]. Therefore, it enhances the generalized linear model by incorporating the flexibility of nonparametric smoother and thus provides additional flexibility and retains the interpretability so important in multipredictor regression analysis. The model can be specified as

$$h(Ey) = \alpha + \sum_i Si(Xi) \qquad (4)$$

where the α is the intercept and Si is the smoother for its corresponding X_i. An important property of smoother is its nonparametric nature, which doesn't assume a rigid form for the dependence of Y on $X_1,\ldots,X_p$ and it produces estimate of the trend that is less variable than Y itself. Frequently used nonparametric smoothers are regression

smoothers such as loess, using running moving regression to predict the value at the center of window and the cubic smoothing splines, which minimizes the penalized residual sum of squares [21].

The estimation of α and Si is accomplished by the local scoring algorithm which replace the weighted linear regression in the adjusted dependent variable regression by an appropriate algorithm for fitting a weighted additive model.

The deviance [22], or likelihood-ratio statistic, for a fitted model $\hat{\mu}$ is defined as $D(y;\hat{\mu}) = 2\{l(\mu_{\max};y) - l(\hat{\mu};y)\}$, where $\mu_{\max}$ is the parameter value that maximizes $l(\mu;y)$ over all μ (the saturated model). The deviance plays the role of the residual sum of squares for generalized models, and in nonparametric and additive models, it can be used for assessing goodness-of-fit and for comparing models.

1.4 Objectives & Hypotheses

Although many studies have investigated the relationship of inorganic air pollutants with asthma exacerbation, few have focused on the association of outdoor pollen and fungal spore concentrations with asthma, especially from a quantitative point of view. Furthermore, even for the studies that were carried out quantitatively, their validity was questioned because of the limitation of their aeroallergen sampling method and unavailable information of individual aeroallergen types.

The objective of this study was to determine the relationship of daily number of hospital visits for treatment of acute pediatric asthma attacks and daily concentration of different types of outdoor pollen and fungal spores in Cincinnati area. The Button Sampler was used to perform the pollen and fungal spore sampling because it has

demonstrated high efficiency over Rotorod Sampler in measuring small bio-particle size [23]. Three hypotheses for this thesis can be summarized as follows:

1. Different types of pollen and fungal spores have different influences on the exacerbation of pediatric asthma.
2. Delayed effect is expected for certain types of pollen and fungal spores in their way of exacerbating pediatric asthma.
3. Both parametric generalized linear model and nonparametric generalized additive model can provide a good fit.

The investigational results from this study will help clinicians in selecting the proper antigenic extract of pollen or fungal spores for skin-prick tests and will provide insightful information to health authorities when initiating strategies for asthma prevention.

Chapter 2. Materials & Methods

2.1 Hospital Asthma Admissions

Cincinnati Children's Hospital Medical Center provided records of daily asthma visits from March 1, 2002 to October 31,2002. An "asthma visit" was defined as either an emergency room visit or outpatient clinic visit for treatment of acute asthma. Records of daily asthma hospital visits were obtained by means of a hospital computer search for all admission with an ICD9 code of any 493.x0 and 493.x1. Patients' age was from 1-18 years old, and infants (0-12 months) were excluded because of the possible confusion of asthma with viral infections. If a person had >1 asthma visits in a day, only 1 was counted.

2.2 Aeroallergen Exposure and Weather Data

Daily sampling of outdoor aeroallergens was performed by investigators in the Center for Health-Related Aerosol Studies at Environmental Health Department. The rooftop of a two-storied office building about 3 miles north of the downtown Cincinnati was selected for sampling. The height of the rooftop was about 7 meters. The nearby vegetation was sparse and there were no tall buildings in the proximity allowing free movement of wind and spatially uniform intensity of the solar radiation at the sampling location. The Button Sampler (SKC, Inc., Eighty Four, PA) is a personal inhalable sampler with a curved porous inlet. It was chosen to perform the sampling because of its demonstrated high efficiency for aeroallergens of small particle size and low sensitivity to the wind direction and velocity [23]. Two Button samplers were fixed onto a sampling tripod about 7.5cm below a rain shield. The position of the Button Samplers was vertical to ground level and they were placed back to back so that the two inlets were oriented

180° opposite to each other: one towards south-west (SW) direction and the other towards north-east (NE).

The sampling began in March 2002 and continued until the end of October 2002. Both Button Samplers were operated at a flow rate of 4L/min continuously for 24 hours from Sunday to Friday starting every morning at about 8:30AM. Thus the pollen and spore counts reported on a given day were collected during the preceding day and night, and the morning of the report day.

When the pollen and fungal spores were collected with the Button Sampler, the count was performed on 40 randomly selected microscopic fields using a Nikon (Labophot 2, Nikon Corp., Japan) high-resolution light microscope. For most samples the magnification of 100X was used for the pollen enumeration, except the cases when the grains were not identifiable or the deposition was too dense (the magnification of 400X was applied in these cases). The fungal spore enumeration was always performed at 400X. Phase contrast objectives were used to identify unpigmented hyaline spores.

The total pollen or fungal spore count, $N_{total\ (BUTTON)}$, on the filter was calculated as follows:

$$N_{total\ (BUTTON)} = N_f \times A_{total\ (BUTTON)}/A_f \qquad (5)$$

where N_f = average number of spores/pollen per microscopic field, $A_{total\ (BUTTON)}$ = total area of the filter (380 mm^2) and A_f = area of a microscopic field (0.1452 mm^2 at 400X magnification and 2.19 mm^2 at 100X magnification).

The flow rate of the Button Sampler operating in the personal sampling mode is 4 L/min. Following the quality assurance protocol, the sampling flow rate was measured by a DryCal® DC-Lite Calibrator (SKC, Inc., Eighty Four, PA) before and after each 24-

hour sampling. The average flow rate, F, from these two measurements, the sampling time, t, and the total pollen/fungal count, $N_{total\ (BUTTON)}$, were used to calculate the total airborne concentration, C_{BUTTON} as follows:

$$C_{BUTTON} = N_{total\ (BUTTON)} / F \times t \qquad (6)$$

Because studies have shown the most common types of allergenic fungal spore are *Alternaria*, *Cladosporium* and the *Aspergillus* species, these three individual species along with the total fungal spore were selected to investigate their association with childhood asthma. Results of a paired t-test [23] showed that for fungal spores, no statistically significant difference was observed between the bio-particle concentrations determined with the two Button Samplers facing opposite directions. Thus daily concentrations of *Aspergilli*, *Cladosporium*, *Alternaria* groups and total fungal spores were obtained from the Button Sampler facing southwest, for the days when southwest data was not available, the same day concentrations from the Button Sampler facing northeast were substituted. With respect to pollen grains, since the types of pollen that most commonly cause allergic reactions are produced by plain-looking plants (trees, grasses, and weeds), statistical analyses would be performed on Ambrosia (Ragweed), Poaceae (Grass), Quarcus (Oak/cedar) and total pollen grains, in order to seek their relationship with childhood asthma. Because the paired t-test for pollen grains showed that the SW Button Sampler had significantly different concentrations from the NE Button Sampler, daily concentrations of Ambrosia (Ragweed), Poaceae (Grass), Quarcus (Oak/cedar) and total pollen grains were obtained by taking the average of measurements gained from SW and NE Button Samplers.

Daily Cincinnati meteorological data from March 1, 2002 to October 31, 2002 was obtained from the website of the National Virtual Data System (http://nndc.noaa.gov), which provided local climatological data with hourly, daily, and monthly summaries for over 700 U.S. locations. Hourly temperature and relative humidity was downloaded with the meteorological information measured at the station at Cincinnati-Northern Kentucky international airport (CVG, WBAN #93814).

In order to keep the consistency in monitoring time with that of the aeroallergen data, daily temperature and relative humidity were calculated by taking the average of 24 continuous hourly observations starting every morning at about 8:30 AM. Thus the temperature and relative humidity reported on a given day were monitored during the preceding day and night, and the morning of the report day.

2.3 Statistical Methods

2.3.1 *Generalized Linear Model*

It is important to remove any known causes of variability in asthma hospital visits prior to examining the association of short-term changes in aeroallergen concentrations with short-term changes in asthma hospitalization. For example, in order to ensure that fluctuations such as the seasonal variations in asthma hospitalization rates do not contribute to the association under investigation, indicator variables for month were created and inserted into the model, and day-of-the-week indicator variables were used to model the variations of daily asthma visits during the week. In order to capture any time trends in asthma hospital visits, a time term was added, and two methods were tried to appropriately model this term, and these include using a linear and quadratic time

term and fitting restricted cubic splines for time term. Method selection would be based on the improvement in model fit.

The dependence of asthma hospitalization on weather may not be linear and a linear term of weather parameters would be an inadequate control, therefore a flexible approach was adopted by which both weather parameters were divided into a number of ranges. Separate dummy variables were created to control for each range of temperature and each range of humidity. This control for the effect of each range of temperature and humidity allows a potentially nonlinear pattern, which is possible in the time series data. The number of categories into which temperature and relative humidity were divided was chosen by starting with five categories, and incrementally increasing the number of categories until the improvement in model fit from an additional category was not statistically significant. Diagnostic plots of residuals of hospital counts were also used to assess the adequacy of the models in controlling for time trend, seasonal variation, and weather dependence. For example, if the model systematically underestimates hospital admission on very cold days, then most of the residuals will be positive on those days. This can be seen on a plot of residuals versus temperature. Nonparametric smoothing was also used to assure no patterns in the residuals, which otherwise would indicate the presence of a poor fit in certain time periods, or in certain weather parameters' ranges. Only after the diagnostics indicated the models controlled for temporal patterns and weather dependence was particular aeroallergen considered.

Time series data often are serially correlated. That is, two days close together in time are likely to be more similar in their hospital admission counts than two randomly chosen days. Although this may be due to the weather, there may be serial correlation

remaining in the residuals of the regression models. Autoregressive models were fit to the residuals to see if any significant autocorrelation existed. If autocorrelation was present, autoregressive Poisson models were estimated using the generalized estimating equations of Liang and Zeger [24].

The Poisson model assumes that the mean and variance are equal. However, many actual count processes are overdispersed which means that the variances are larger than expected. Overdispersion tends to cause underestimation of standard errors of the regression coefficients, thus leads to overestimated test statistics and consequently to biased hypothesis tests. Overdispersion can be modeled by letting the actual variance equal the assumed variance multiplied by an additional parameter that adjusts for the discrepancy between assumed and actual. For the Poisson distribution, the assumed variance is the mean, λ, so the adjusted variance is $\phi\,\lambda$, and ϕ is typically referred to as the overdispersion parameter. McCullagh and Nelder [20] suggested using the deviance to estimate the overdispersion parameter. That is $\widehat{\phi} = \dfrac{Deviance}{N-P}$.

2.3.2 *Generalized Additive Model*

The association of short-term changes in aeroallergen concentrations with short-term changes in asthma hospitalization was investigated using generalized additive Poisson regression models. Assuming linear aeroallergen and hospital admission association, the model is of the form:

$$\begin{aligned}\log E(\textbf{asthma hospital visits}) = \alpha + \beta\,\textbf{aeroallergen} + \mathbf{S}_1(time) + S_2(temperature) \\ + S_3(relativehumidity) + day-of-week\ \textbf{indicators}\end{aligned} \qquad (7)$$

where S_1, S_2 and S_3 are the nonparametric smoothers respectively for time, temperature and relative humidity. Time is used as a proxy for any outcome predictors not included

in the model, which have long-term trends or seasonal patterns. Hence long-term trends and seasonal patterns from the data are removed with a smooth function of time to guard against the confounding by omitted variables. Smoothing splines was the nonparametric smoother chosen to fit the model and this approach involves the choice of a smoothness parameter, which equals to $tr(s)-1$, where $tr(s)$ is the trace of implicit smoother matrix. The selection approach taken here was to start with the default smoothing parameter, and then increase it incrementally until no significant improvement in both the model fit and appearance of the diagnostic plots.

Chapter3. Results

3.1 Hospital Asthma Admissions

For the months of March through October in 2002, there were 1950 cases of childhood asthma admitted to CHMCC's emergency room or outpatient clinic. Among these cases, 29 were repeated hospital visits that occurred for the same patient on the same day, thus these cases were excluded from the subsequent analyses. Examination of daily asthma hospital visits (Figure 1) showed that the highest number of visits occurred in March, May, September and October. Monthly variation in hospital asthma visits was further examined by tabulating monthly means (Table 1).

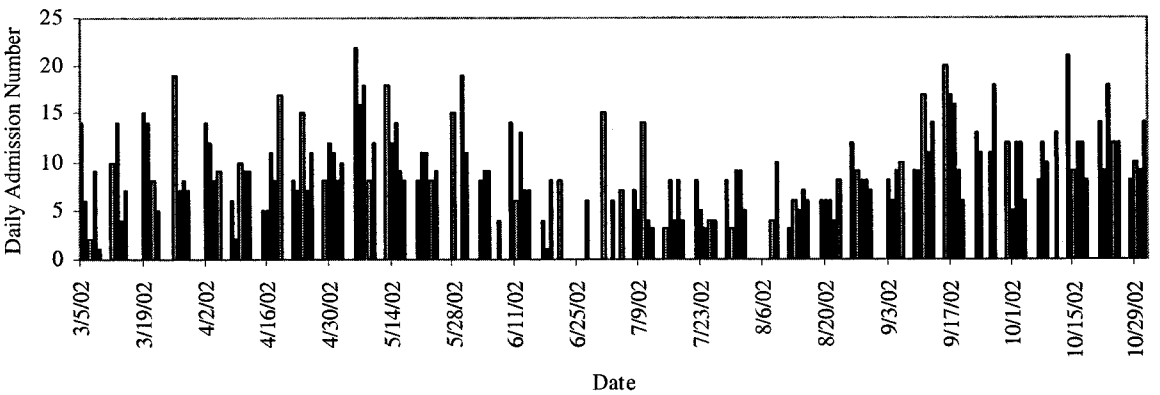


Figure 1.Daily hospital asthma admission from March to October 2002.

Table 1. Means of daily hospital admission (± standard deviation) from March to October 2002

	March	April	May	June
Visits/Day	8.8 (4.9)	9.2 (3.5)	12.3 (4.2)	7.4 (3.4)

	July	August	September	October
Visits/Day	6.3 (3.4)	6.8 (2.2)	11.9 (4.1)	11.2 (3.6)

3.2 Aeroallergen Exposure and Meteorological data

Outdoor aeroallergen sampling was performed daily during the weekdays for the months of March through October in 2002. Excluding the unreliable observations due to the conditions of bad weather or broken sampler, there were 154 daily records available with concentrations of aeroallergen types that we were intended to investigate.

The types of pollen that most commonly cause allergic reactions are produced by the plain-looking plants (tress, grasses, and weeds), and their seasonal patterns differ dramatically. Ragweed pollen pollinated from August to October, with a peak presence in September, while for Oak/Cedar pollen, its prevalence was mainly in spring from March to May. Grass pollen began to appear in April, and although its outdoor appearance was throughout the rest of the whole year, its greatest concentration occurred in the beginning of June. Because of the vast varieties of pollen grains and their different pollination seasons, pollen grains were prevalent through out the whole year with the peak total pollen concentration appearing in April.

Large numbers of fungal spores are usually present in outdoor air and frequently exceed pollen concentration by 100-to 1000-fold. *Alternaria, Cladosporium* and the *Aspergillus* species are the most common and most allergenic types of fungal spores. Although these were found through the whole year, their peak presence was in October, similar to the seasonal pattern for total fungal spore concentration. Figures 2 and 3 illustrate seasonal variations for pollen grains and fungal spores.

The monthly mean temperature and mean humidity ranged from 43.8 to 77.9°F and from 66.4 to 75.8, respectively. Correlations between aeroallergens and meteorological variables are shown in Table 2 and 3, their monthly means are presented in Table 4.

Table 2. Correlation Coefficients for Specified Pollen Grains and Temperature, Humidity

	Temp	Humid	Grass	Ragweed	Oak/Cedar	Total
Temp	1	-0.04	0.23	0.28	-0.08	0.07
Humid		1	-0.03	-0.22	-0.02	-0.04
Grass			1	-0.06	-0.07	-0.01
Ragweed				1	-0.11	-0.04
Oak/Cedar					1	0.50
Total						1

Note. Temp, temperature; Humid, humidity; Total, total pollen grains.

Table 3. Correlation Coefficients for Specified Fungal Spores and Temperature, Humidity

	Temp	Humid	*Alternaria*	*Cladosporium*	*Aspergilli*	Total
Temp	1	0.02	-0.21	-0.01	0.01	-0.07
Humid		1	0.04	0.12	0.20	0.19
Alternaria			1	0.67	0.59	0.75
Cladosporium				1	0.72	0.89
Aspergilli					1	0.93
Total						1

Note. Temp, temperature; Humid, humidity; Total, total fungal spores.

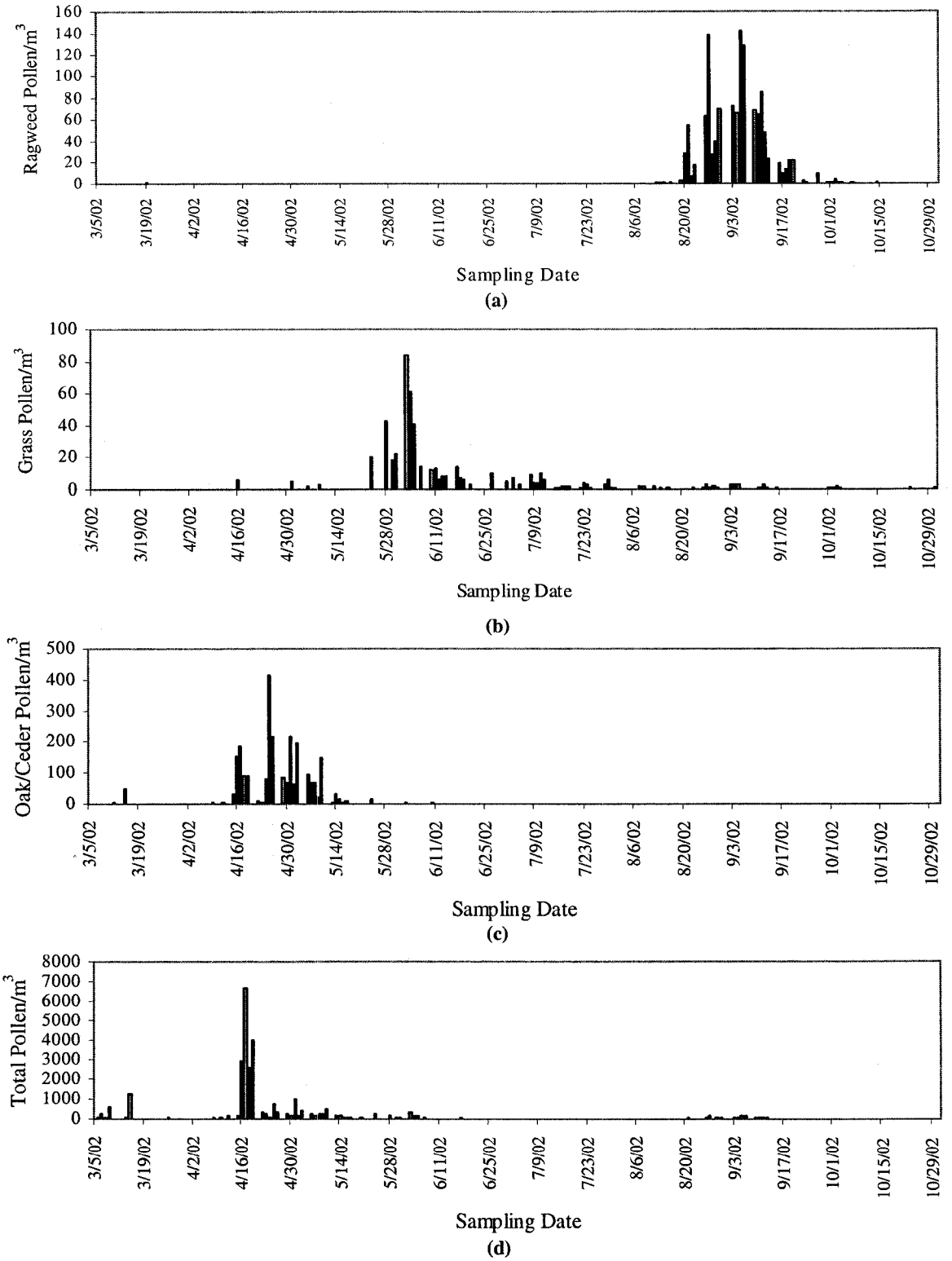
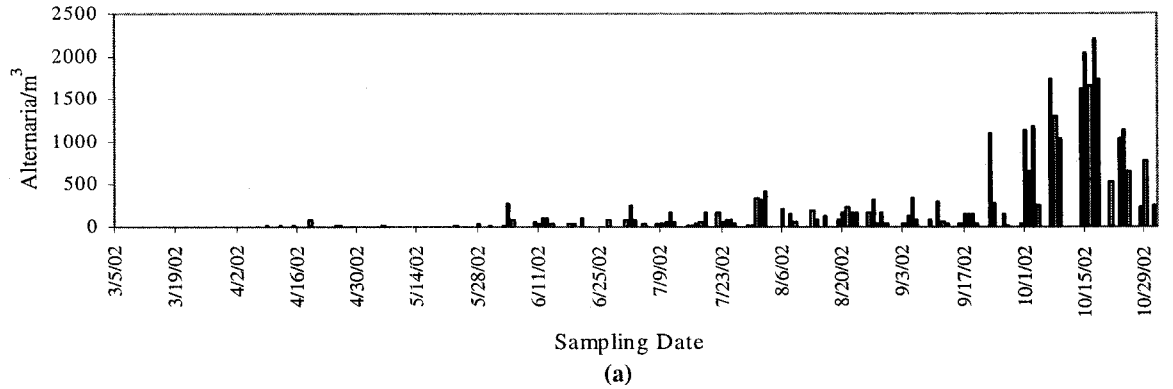
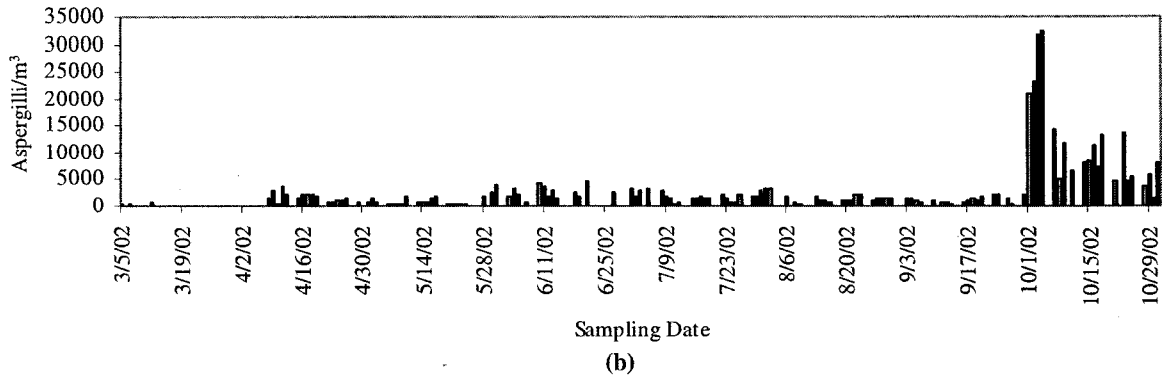


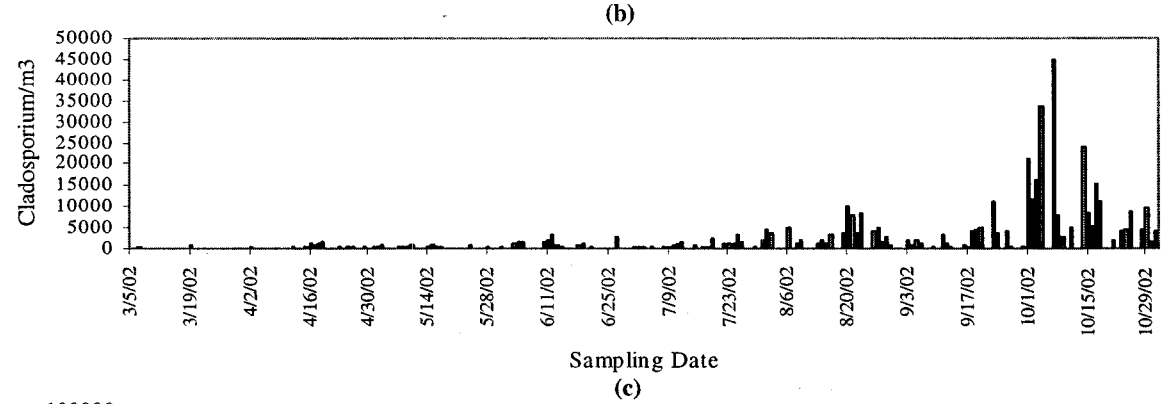
Figure 2. Daily outdoor pollen grains: (a) Daily Ragweed pollen concentration; (b) Daily Grass pollen concentration; (c) Daily Oak/Cedar pollen concentration; (d) Daily Total pollen grain concentration



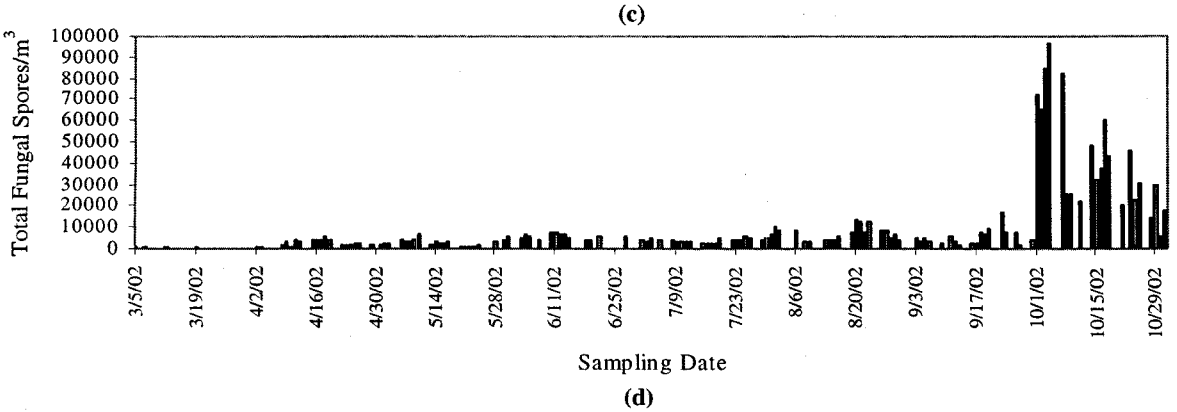
(a)



(b)



(c)



(d)

Figure 3. Daily outdoor fungal spores: (a) Daily *Alternaria* concentration; (b) Daily *Aspergilli* concentration; (c) Daily *Cladosporium* concentration; (d) Daily total fungal spore concentration

Table 4. Monthly Means of Daily Aeroallergen Concentration (± Standard Deviation) from March to October 2002

Aeroallergen	March	April	May	June
Ragweed Pollen /m^3	0.0	0.0	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)
Grass Pollen /m^3	0.02	0.3	5.5	20.5
	(0.09)	(1.2)	(11.2)	(24.1)
Oak/Cedar Pollen /m^3	3.3	69.1	46.2	0.8
	(11.6)	(103.5)	(66.4)	(1.1)
Total Pollen /m^3	147.9	903.1	202.8	60.9
	(322.5)	(1732.5)	(224.5)	(79.6)
Alternaria /m^3	1.9	8.4	3.2	70.6
	(3.8)	(18.2)	(8.1)	(67.8)
Aspergilli /m^3	236.9	1367.1	1043.8	2467.2
	(265.0)	(975.7)	(951.1)	(1311.0)
Cladosporium /m^3	171.4	395.3	412.0	1338.7
	(251.6)	(482.1)	(228.1)	(1008.5)
Total Fungal /m^3	574.1	2557.3	2861.2	5486.7
	(405.6)	(1646.5)	(1559.4)	(1918.6)

Aeroallergen	July	August	September	October
Ragweed Pollen /m^3	0.0	23.9	41.9	0.5
	(0.0)	(36.2)	(43.0)	(2.8)
Grass Pollen /m^3	3.6	1.1	1.0	0.4
	(5.6)	(36.1)	(43.0)	(1.0)
Oak/Cedar Pollen /m^3	0.0	0.0	0.0	0.0
	(0.0)	(0.0)	(0.0)	(0.0)
Total Pollen /m^3	6.3	26.5	47.2	2.1
	(4.5)	(36.9)	(44.7)	(2.8)
Alternaria /m^3	88.2	160.6	172.7	961.6
	(80.8)	(105.0)	(247.5)	(687.5)
Aspergilli /m^3	1848.9	1515.3	1239.3	11048.9
	(805.9)	(796.7)	(615.1)	(8861.5)
Cladosporium /m^3	976.1	3841.6	2596.5	11281.9
	(816.3)	(2566.6)	(2728.7)	(11134.4)
Total Fungal /m^3	3986.4	7248.7	5400.6	40149.4
	(1227.0)	(3265.1)	(3739.6)	(26538.1)

3.3 Statistical Analyses

3.3.1 Generalized Linear Model

The fraction of the population visiting a hospital for asthma on a given day is quite small, thus Poisson regression model was used to examine the dependence of asthma hospital visits on aeroallergens. SAS Version 8e was the statistical software used for all analyses in this session. One big challenge occurred in daily hospital admission study is the substantial seasonal and other temporal patterns, it is critical to ensure these patterns do not contribute to the association between aeroallergen concentration and daily asthma hospital visits. This problem was addressed by controlling monthly variations with dummy variables for each of the 8 studying months and by including a linear and quadratic time term to capture any secular trends in asthma hospital visits. However, many time series count data including the one of this study are overdispersed (the variance is greater than the mean) because of the uncontrolled differences in the underlying risk (social economic status, health care plan, etc) of different subpopulations, so for all the subsequent statistical analyses in this section, overdispersion was modeled by specifying option SCALE=P in Proc GENMOD, MODEL statement. This would allow SAS to assume the overdispersion parameter to be given by Pearson's chi-square statistic divided by its own degrees of freedom and allow all statistics such as standard errors and likelihood ratio statistics adjusted appropriately. Although including linear and quadratic time terms is a standard way to capture continuous temporal trend, it could be seen from figure 4 that it may not be an effective method for this study, and a localized regression model are preferred. Therefore, the restricted cubic spline function (also called natural splines) was chosen to control temporal pattern. Its advantages over cubic spline

function, which is a popular choice of regression splines, are its good tail fitting by constraining the function to be linear in the tails and the fewer number of parameters to be estimated [25]. Five knots of time were selected with their positions at the quintiles of 0.05, 0.275, 0.50, 0.725, 0.95, respectively. The restricted spline function is given by

$$F(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{k-1} X_{k-1} \qquad (8)$$

where X1=Time and for J=1……, k-2,

$$X_{j+1} = (X - t_j)^3_+ - (X - t_{k-1})^3_+ (t_k - t_j)/(t_k - t_{k-1}) + (X - t_k)^3_+ (t_{k-1} - t_j)/(t_k - t_{k-1}).[25]$$

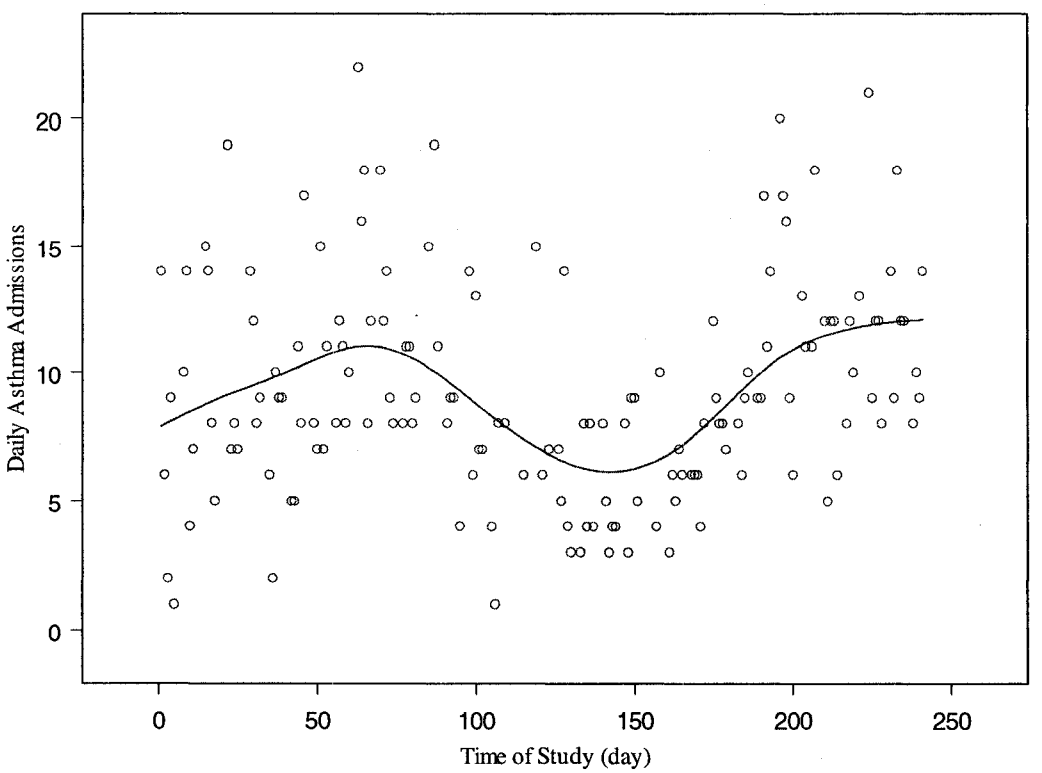


Figure 4. Plot of daily asthma hospital visits over study time from March to October in 2002. It shows a long-term temporal pattern.

Results shown in table 5 were obtained from the two different methods in capturing secular trends, one used the linear and quadratic time terms and the other used the restricted cubic splines. Deviance in generalized linear model is defined as

$2[l(\theta(y);y)-l(\theta(X\hat{\beta});y)]$ [20], where $l(\theta(y);y)$ is the log likelihood, with $\theta(y)$ the value of θ determined from the data and $\theta(X\hat{\beta})$ determined from the estimate of β under the model. The deviance has an approximate χ^2 distribution with *N-P* degrees of freedom, where *N* is the number of observations and *P* is the rank of *X*. For distributions that do not depend on an unknown scale parameter, such as the Poisson, the deviance tests lack of fit of model. From table 5, we could see that for the model with a linear and quadratic time term, the deviance was 285.2 with 151 degrees of freedom, while for the model with restricted cubic splines, the deviance decreased to 234.4 with 149 degrees of freedom. Therefore, due to the significant improvement in model fit, restricted cubic splines were chosen to control temporal trend.

Table 5.Results of Two Models with Different Methods in Controlling Temporal Trend

Model	Effects	Estimates	Standard Error	Chi-Square	Pr>Chi
LogE (daily asthma visit)=Time Time*Time Deviance=285.2, DF=151	Intercept	2.39	0.11	534.0	<0.001
	Time	-0.01	2.00E-3	6.3	0.01
	Time*Time	2.31E-5	8.08E-6	8.2	<0.01
LogE (daily asthma visit)=Restricted cubic spline terms of Time Deviance=234.4, DF=149 Note: Time were transformed into variables X1-X4	Intercept	1.91	0.14	191.2	<0.001
	X1	0.01	3.6E-3	13.8	<0.001
	X2	-1.97E-6	4.14E-7	22.6	<0.001
	X3	5.46E-6	1.16E-6	22.1	<0.001
	X4	-6.31E-6	1.68E-6	14.2	<0.001

After dummy variables for month were added thereafter, the deviance for the model controlling long-term temporal and seasonal patterns became 210 with 142 degrees of freedom, and this improvement in model fit (P=0.001) indicated significant month effect on daily number of hospital asthma visits.

Asthma attacks may be triggered directly by weather conditions, or weather may be associated with infectious disease that subsequently triggers attacks [26]. Therefore, its effect needs to be filtered and only after the model controlled for temporal patterns and weather were aeroallergen influences considered. Because the exact shape of the dependence on temperature and humidity is unknown and it may not be linear, they were both divided up into a number of ranges set upon by percentiles of equal distance, and dummy variables were created for each range. The number of categories into which temperature and humidity were divided was chosen by starting with five categories, and incrementally increasing the number of categories. Univariate Poisson regression was performed at each step and the selection of category number was based on the degree of improvement in model fit. Table 6 and 7 are the statistics generated from the category selection procedure for temperature and humidity, and they were used to explore the ways to control for weather.

A type I analysis, which is sometimes called an analysis of deviance, consists of fitting a sequence of models, beginning with a simple model with only an intercept term, and continuing through a model of fitting one additional effect on each step. For the dispersion parameter held fixed for all models, which is the case for this study, likelihood ratio statistics defined as twice the difference of the log likelihood between successive models is equivalent to computing the scaled difference of deviances. The asymptotic

distribution of the likelihood ratio statistics, under the hypothesis that the additional parameter included in the model are equal to 0, is a chi-square with degrees of freedom equal to the difference in the number of parameters estimated in the successive models, and improvement of model fit can also be based on this statistics. From table 6, we could see that temperature with 6 categories was best to control for weather dependence, because it lead to the most significant improvement in model fit. While temperature was an influential factor for asthma hospital visits, humidity didn't seem to be so according from the results in Table 7, thus it was excluded from the subsequent analyses.

Table 6. Statistics of Analysis of Deviance (Type I Analysis) of the Selection of Temperature Categories

Model	Deviance (DF)	Scaled Deviance Difference	Pr > ChiSq
Intercept only	302.41 (153)	-	-
LogE(Y)=Temperature, 5 categories	278.63 (149)	12.90	0.01
LogE(Y)=Temperature, 6 categories	249.35 (148)	28.57	<0.0001
LogE(Y)=Temperatue, 7 categories	276.01 (147)	14.17	0.02
LogE(Y)=Temperature, 8 categories	271.31 (146)	16.85	0.02
LogE(Y)=Temperature, 9 categories	273.96 (145)	15.23	0.05

Note: The scaled deviance differences refer to the scaled deviance difference between tested model and intercept only model.

Table 7. Statistics of Analysis of Deviance (Type I Analysis) of the Selection of Humidity Categories

Model	Deviance (DF)	Scaled Deviance Difference	Pr >ChiSq
Intercept only	302.41 (153)	–	–
LogE(Y)=Humidity, 5 categories	293.85 (149)	4.41	0.35
LogE (Y)=Humidity, 6 categories	290.24 (148)	6.57	0.25
LogE (Y)=Humidity, 7 categories	286.48 (147)	8.26	0.22
LogE(Y)=Humidity, 8 categories	286.42 (146)	8.18	0.32
LogE(Y)=Humidity, 9 categories	286.95 (145)	7.94	0.44

Note: The scaled deviance differences refer to the scaled deviance difference between tested model and intercept only model.

It is commonly believed that number of hospital visits varies through the week with possible greater numbers on Monday. Therefore, day of the week indicator variables were created and added to the model no matter model fit improvement was significant or not. Diagnostic plots of the residuals of hospital counts were generated to assess the adequacy of the model in controlling for trend, seasonal variation, and weather dependence. Figure5 shows the residuals from the regression model without aeroallergens. The seasonal pattern has been removed, as illustrated by the nonparametric smooth. Similar diagnostic plot (Figure 6) was examined with residuals versus temperature divided into 6 ranges and nonparametric smooth indicates a good control for temperature dependence.

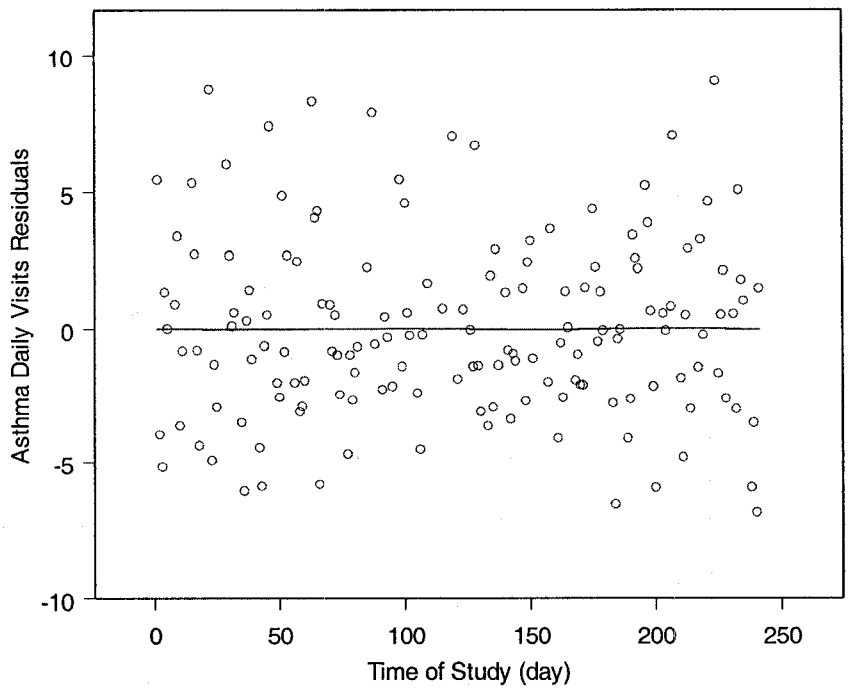


Figure 5. Residual of daily hospital visits versus time of study during the study period. Nonparametric smooth line indicates good control for temporal pattern.

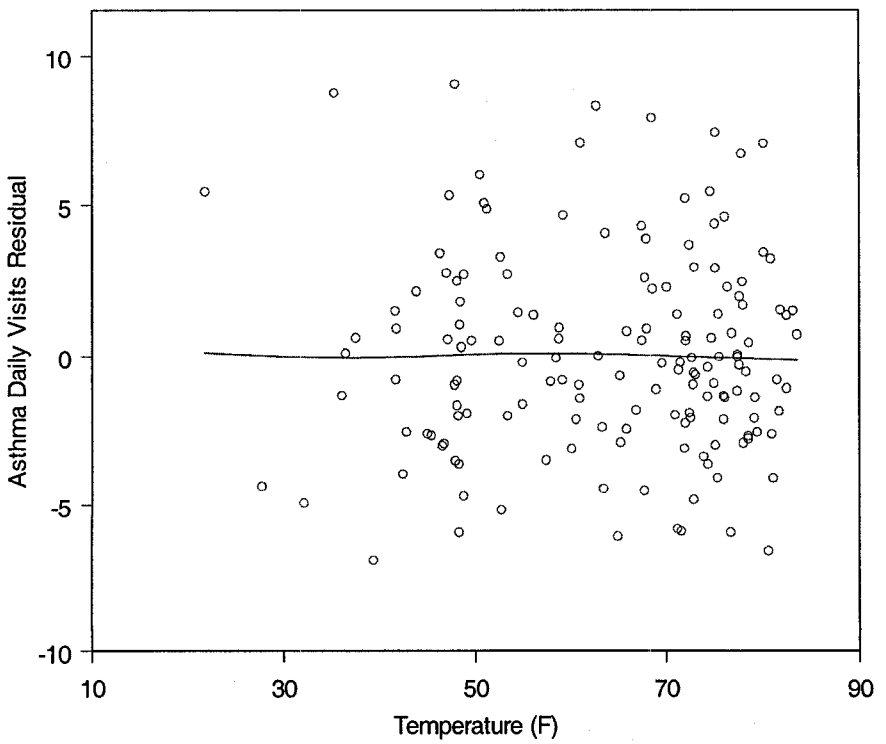


Figure 6. Residual of daily hospital visits versus daily temperature during the study period. Nonparametric smooth line indicates good control for temperature dependence.

A linear term of aeroallergens was added to the model after trend, seasonal variation, and weather dependence have been controlled. Each aeroallergen type was considered singly, and only those presented throughout the whole study period, were added. For Ragweed pollen and Oak/Cedar pollen, different models controlling for temporal trend and weather dependence were fitted with data of their appearing months only. Meanwhile, since the amount of outdoor aeroallergens may not result in an immediate asthma visit, aeroallergen concentrations from 0 to 5 days before the asthma visit date were considered. Among different types of aeroallergens with various lag days, total pollen concentration 2 days before the visit and *Cladosporium* concentration 4 days before the visit were significant predictors for asthma visits.

For Ragweed pollen, since its outdoor existence was only through August to October in 2002, analyses were performed on the data of these 3 months, and so were for Oak/Cedar pollen with analyses on the data of March through June. Because of the smaller sample size, 3 knots for time were selected for restricted cubic spline function at the percentiles of 0.1, 0.5 and 0.9. Same methods for temperature category selection were applied, and temperature was chosen to be divided into 3 and 6 categories, respectively for Ragweed pollen and Oak/Cedar pollen. Concentrations of aeroallergens with 0-5 lag days were entered into the model after it has been controlled for seasonal pattern and weather dependence. It was found that Oak/Cedar concentration 3 days before the visit was a significant asthma predictor.

Autoregressive models were fit to the residuals, and no significant serial correlation was found. All the significant results are summarized in Table 8.

Table 8. Statistics of Significant Aeroallergen Predictors from Poisson Regression Model

Aeroallergen (lag days)	Coefficient Estimate	Standard Error	Relative Risk	95% CI
Total Pollen (2)	8.30E-5	4.08E-4	1.008	1.0003-1.016
Cladosporium (4)	2.44E-5	7.77E-6	1.002	1.001-1.004
Oak/Cedar Pollen (3)	2.7E-3	1.3E-3	1.31	1.041-1.649

Note: The relative risk is for a 100 counts/m^3 increase in aeroallergen concentration.

3.3.2 *Generalized Additive Model*

Generalized additive model is a nonparametric method in which it enhances the generalized linear model by incorporating the flexibility of nonparametric smoother and retain the interpretability. It is an appropriate model for time series environment study when little knowledge is known about the nature of correlation. The approach taken here was to fit the counts of daily asthma visits to nonparametric smoothed functions of time, temperature and linear terms of aeroallergen concentration. Time was used as a proxy for any outcome predictors not included in the model, which have long-term trends and seasonal patterns. Thus these patterns were removed with a smooth function of time to guard against the confounding by omitted variables. Degrees of freedom were saved because dummy variables for each month were not included as they were in the generalized linear model. Smoothing spline was the nonparametric smoother chosen in this study and the choice of its smoothing parameter was critical because it determined how smooth the curves should be. S-Plus 6.1 and SAS Version 8e were the statistical packages used to fit generalized additive models, where smoothing spline is defined as S (X, df=a), here X is the univariate predictor and df is the target equivalent degrees of freedom, used as a smoothing parameter, its value equals to tr(S)-1, where S is the

implicit smoother matrix and tr(S) is its trace. The approach taken here was to start with the default smoothing parameter (df=4), and then increase it until residual plots showed adequate control and no significant improvement in model fit. In generalized additive model, for a fitted model $\hat{\mu}$, deviance is defined as $D(y;\hat{\mu}) = 2\{l(\mu_{\max}; y) - l(\hat{\mu}; y)\}$, where $\mu_{\max}$ is the parameter value that maximizes $l(\mu; y)$ over all μ (the saturated model). The deviance plays the role of the residual sum of squares for generalized models, and in nonparametric and additive modes, it can be used for assessing goodness-of-fit and for comparing models. Meanwhile, suppose η_1 and η_2 are two linear models, with η_1 nested within η_2. If η_1 is assumed to be correct, then $D(\hat{\eta}_2;\hat{\eta}_1) = D(y;\hat{\eta}_1) - D(y;\hat{\eta}_1,\hat{\eta}_2)$ has an asymptotic $\chi 2$ distribution with degrees of freedom equal to the difference in the dimensions of the two models. This result is used extensively for comparing models, and often presented in the form of an analysis of deviance table. Smoothing spline terms of Time and Temperature were fitted separately, and Table 9 and 10 showed deviances generated from the procedures of their smoothing parameter selection. From Table 9, we could see model fit was improved significantly with the increasing degrees of freedom until it reached 8. However, from diagnostic plot (Figure 7), time smoothing spline with 9 degrees of freedom seemed to have a better control for temporal pattern than that with 8 degrees of freedom, thus smoothing spline of Time with 9 degrees of freedom was selected. After the same selection procedure was applied to Temperature, 6 degrees of freedom was selected for its smoothing spline. Figure 8 showed a better control for temperature dependence with smoothing splines of 6 degrees of freedom than that of 5. Therefore, smoothing spline of 9 degrees of freedom for time and 6 degrees of freedom for temperature were adequate in controlling the seasonal pattern and weather

dependence. After dummy variables for day of the week were added to the model, terms of aeroallergen concentration from 0 to 5 days before the visit were entered singly to explore the association of asthma hospital visits with outdoor aeroallergen concentration. Because Ragweed pollen and Oak/Cedar pollen had their unique seasonal appearances, separate models were fitted. Based on the previous criterion for smoothing parameter selection, 6 degrees of freedom was chosen for time smoothing splines for both aeroallergens, while 4 and 6 degrees of freedom were selected for temperature smoothing spline, respectively for Ragweed and Oak/Cedar Pollen.

Table 9. Analysis of Deviance Table for the Smoothing Parameter Selection of Time

Model	Deviance (DF)	Deviance Difference	Pr > ChiSq
Intercept only	302.41 (153)	_	_
Admission= S (time,df=4)	232.55 (149)	69.86	<0.0001
Admission= S (time, df=5)	223.98 (148)	8.57	0.003
Admission= S (time, df=6)	216.12 (147)	7.82	0.005
Admission= S (time, df=7)	210.75 (146)	5.37	0.02
Admission= S (time, df=8)	204.84 (145)	5.91	0.01
Admission= S (time, df=9)	202.14 (144)	2.7	0.10

Note: Deviance difference denotes the deviance difference between two adjacent models.

Table 10. Analysis of Deviance Table for the Smoothing Parameter Selection of Temperature

Model	Deviance (DF)	Deviance Difference	Pr > ChiSq
Intercept only	302.41 (153)	_	_
Admission= S (Temperature,df=4)	276.72 (149)	25.69	<0.0001
Admission= S (Temperature, df=5)	270.48 (148)	6.24	0.01
Admission= S (Temperature, df=6)	268.07 (147)	2.41	0.12

Note: Deviance difference denotes the deviance difference between two adjacent models.

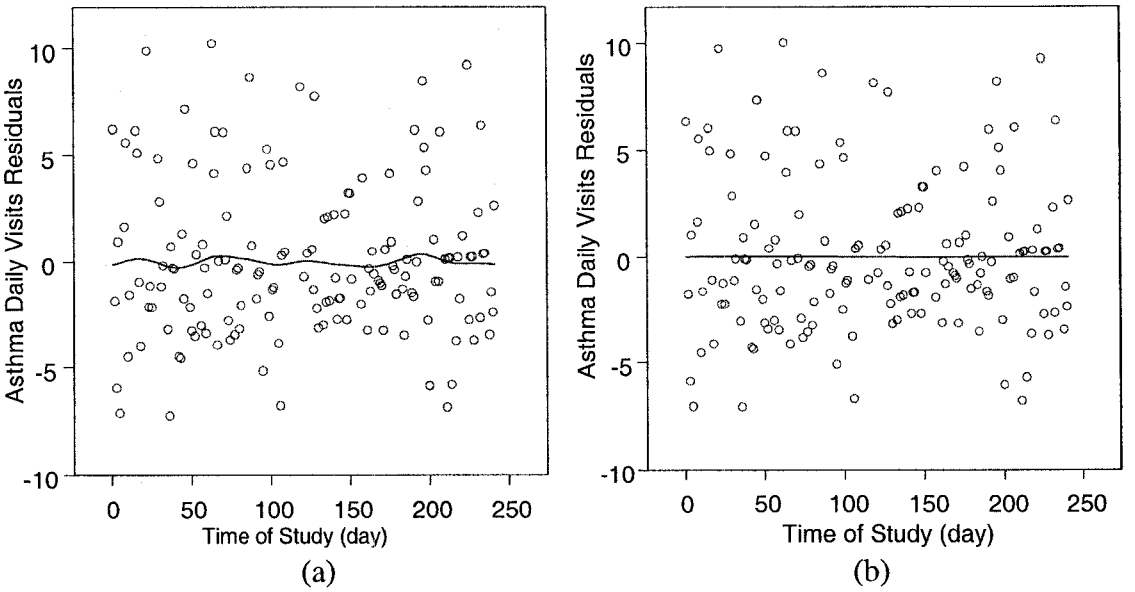


Figure 7. Residual number of daily hospital visits versus study time: (a) residuals obtained after smoothing splines of 8 degrees of freedom was fitted, nonparametric smoothing showed slight evidence of curvature, suggesting an inadequate fit in some regions; (b) residuals obtained after smoothing splines of 9 degrees of freedom was fitted, nonparametric smoothing showed adequate control.

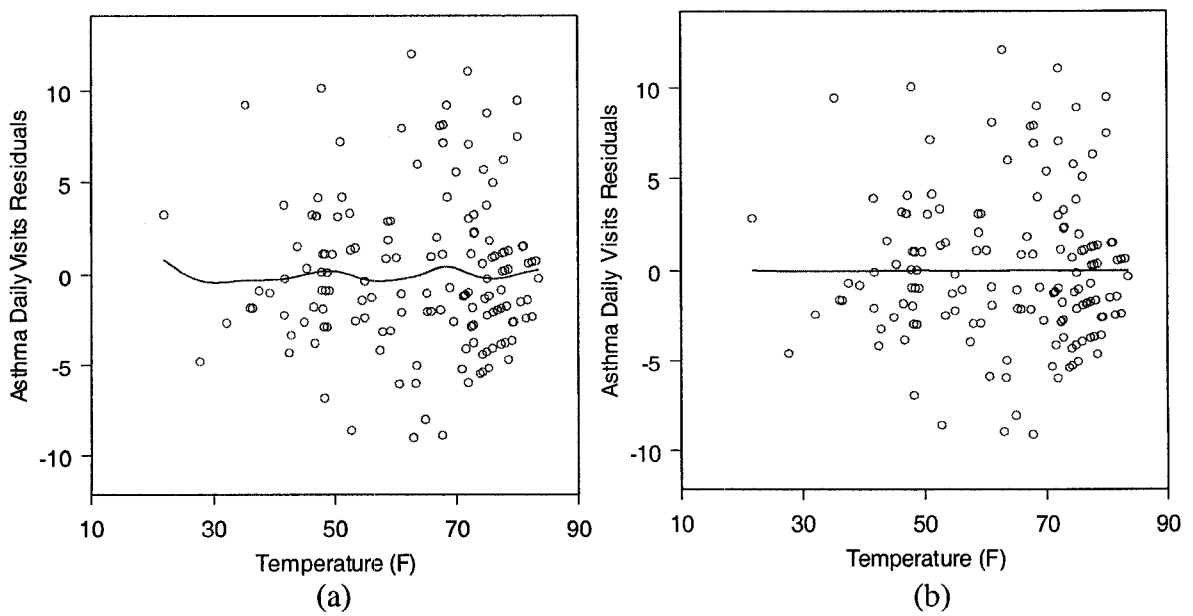


Figure 8. Residual number of daily hospital visits versus daily temperature: (a) residuals obtained after smoothing splins of 5 degrees of freedom was fitted, nonparametric smoothing showed slight evidence of curvature, suggesting an inadequate fit in some regions; (b) residuals obtained after smoothing splines of 6 degrees of freedom was fitted, nonparametric smoothing showed adequate control.

Total pollen concentration 2 days before the visit, Oak/Cedar pollen concentration 3 days before the visit and *Cladosporium* concentration 4 days before the visit were still found to be significant asthma predictors in this model. Their estimated effects were close to those obtained from generalized linear model except that for *Cladosporium*. Meanwhile, since additive model used fewer degrees of freedom and its flexible approach was successful in detecting and controlling for nonlinearities of seasonal trend and weather dependence, more types of aeroallergens were found to be significant. These were Ragweed pollen concentration 4 days before the visit and *Alternaria* concentration 5 days before the visit. All the significant results are summarized in Table 11.

Table 11. Statistics of significant aeroallergen predictors from generalized additive model

Aeroallergen (lag days)	Coefficient Estimate	Standard Error	Relative Risk	95% CI
Total Pollen (2)	8.35E-5	3.84E-5	1.008	1.0006-1.0161
Cladosporium (4)	1.29E-5	5.69E-6	1.001	1.0001-1.0024
Oak/Cedar Pollen (3)	2.1E-3	9.18E-4	1.233	1.0173-1.4961
Ragweed Pollen (5)	4.33E-3	1.97E-3	1.542	1.0193-2.3323
Alternaria (5)	2.88E-4	1.14E-4	1.029	1.0059-1.0530

Note: The relative risk is for a 100 counts/m^3 increase in aeroallergen concentration.

To determine whether the relationship exhibited a linear increase, plots were generated with the relative risk of asthma hospital visits by quartiles of these significant predictors, with the risk in the lowest quartile taken as 1.0. Relative risks were adjusted for temperature, time trends and day of week. For all of these aeroallergens, Figure 9 shows clear evidence of their monotonic increases and no evidence of a threshold.

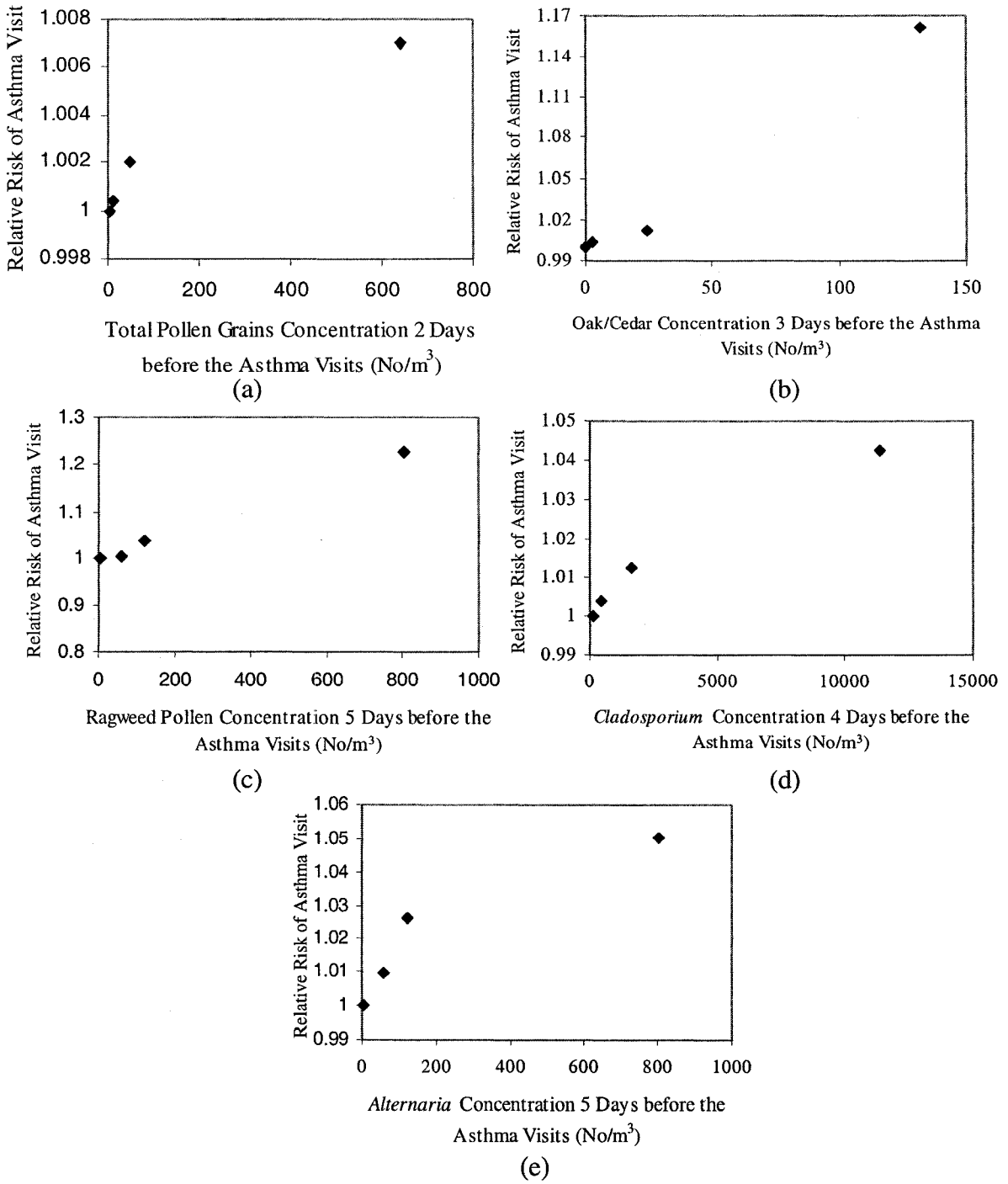


Figure 9. Relative risk of asthma visits by quartile of aeroallergen concentration vs. the mean aeroallergen concentration in the quartile: Relative risk of asthma visit vs. (a) total pollen concentration 2 days, (b) Oak/Cedar pollen concentration 3 days, (c) Ragweed pollen concentration 5 days, (d) *Cladosporium* concentration 4 days, and (e) *Alternaria* concentration 5 days before the visit, respectively.

Chapter 4. Conclusions and Discussion

Based on the modeling and analyses in Chapter 3, the following conclusions were reached.

- Oak/Cedar pollen, Ragweed pollen and total outdoor pollen grains were observed to be significant asthma predictors. Significant results were also found for *Cladosporium* and *Alternaria* groups of fungal spores.
- Different aeroallergens were found to have different delayed effects on asthma sensitization. The lag effect of total outdoor pollen was 2 days, while for Oak/Cedar pollen and Ragweed pollen, the lag effects were 3 and 5 days, respectively. The delayed effects for *Alternaria* and *Cladosporium* groups of fungal spores were 4 and 5 days, respectively.
- The association between asthma admissions and aeroallergen concentrations showed a linear increasing relationship and no evidence of a threshold.
- Both the generalized linear model and the generalized additive model provided a good fit for all in which significance was detected. The estimated coefficients were similar for the significant predictors found in both models. However, more significant results were observed in the latter, due to reduction in degrees of freedom and the more flexible modeling approach.

Asthma is defined as a chronic inflammatory pulmonary disorder, and the airway inflammation noted in asthma is due to an immune-mediated process in which inflammatory cells and inflammatory mediators enter airway tissues to cause disease. The aeroallergens relevant to asthma are the ones that deposit at the lower airways. The site of respiratory deposition of particles relates largely to their aerodynamic diameter,

defined as the diameter of a unit density spherical particle having the same settling velocity as the actual particle. Particles of relatively large aerodynamic diameter of >10 μm (most pollens) are deposited mainly in the nose, pharynx, and upper airway, thus lead to rhinitis; particles of < 10 μm, especially those of < 5 μm are deposited mainly in smaller airways. The aerodynamic diameter term compensates for the fact that particle settling relates not only to actual size but also to shape and density. This helps to explain the significant effects of *Alternaria*, which has a large morphological size (80-100 μm in length) yet a small aerodynamic size (<10μm), and *Cladosporium*, whose morphological size (<10μm in length) and aerodynamic size (<5μm) are both small [27].

Factors other than particle size and concentration also determine inhaled particle deposition. Rhinitis or anatomic nasal abnormalities can lead to mouth-breathing, with the loss of the nasal filter, and thus to more lower airway deposition. This may be one reason that some of the pollen grains found to be related with asthma symptoms. Meanwhile, recent studies have shown particles smaller than intact pollen grain may carry the allergenic activity originating from ragweed pollen, birch pollen, and asthma symptoms can be triggered by the penetrating capability of these small particles into the lower airways [28]. It has also been hypothesized that pollutants can influence the way a pollen, once inhaled, is processed, airway mucosal damage and impaired mucociliary clearance induced by air pollution may facilitate the access of inhaled pollens to the cells of the immune system [29]. All the above factors may contribute to the significant influences on asthma for certain types of pollens and their delayed effects.

Given the method of analysis used, it is unlikely that the found associations are due to unmeasured confounding factors, since the present analysis filters out seasonal and

other long-term temporal trends. Moreover, because daily changes in pollen and fungal concentration were compared to daily changes in asthma visits, the major risk factors for asthma such as socioeconomic and demographic characteristics, would not be expected to confound our findings as they do not change day to day in concert with aeroallergen levels. Aeroallergen sampling with Button Personal Inhalable Aerosol Sampler also adds more creditability to this study, because Button Sampler is more efficient for sampling of outdoor aeroallergens especially the ones of small particle size. In all, this study suggests that routine monitoring of environmental aeroallergens may be a useful indicator of the likelihood of asthma attacks among children and may allow more efficient use of medical resources.

Results of this study were based on 8 months only, so further work is needed to determine whether these observations are typical of all years. In the literatures, PM_{10}, sulfur dioxide, ozon and TSP are the air pollutants found to be significantly related to asthma. Although their relative risks on asthma are close to the ones of the observed aeroallergens in this study, their significances were not directly compared because of their different sampling units. In the future study, relative risks calculated from the standardized datasets of these air pollutants and aeroallergens should be compared, in that case, the comparison would not be affected by the scales and would be more meaningful. Meanwhile, outdoor pollen and fungal concentrations were used in this study to evaluate aeroallergen exposure, but considering the greater amount of time children spent indoors, air sampling of indoor environmental may be pursued in future research. Particulate matter and diesel exhaust may be considered along with aeroallergens in the future asthma study, because it has been demonstrated that a

synergistic interaction exists between particulate, diesel exhaust and pollen exposures in enhancing airway immunoglobulin (Ig)E production and thus lead to severe asthma symptoms [30-32].

Reference

1. Marwick, C. Helping city children control asthma. *JAMA 1997;277:1503-1504.*
2. Skolnick, A.A. Screening to help asthmatics breath easier. *JAMA 1997; 277: 1921.*
3. Adams, P.F., and Benson, V. Current estimates from the National Health Interview Survey. *Vital Health Stat 1990;181,10.*
4. Evans, R., Mullally, D.I., and Wilson, R.W. National trends in the morbidity and mortality of asthma in the US. *Chest 1987; 9:65s-74s.*
5. Barnes, P.J. Air pollution and asthma. *Postgrad. Med. J. 1994; 70:319-325.*
6. Wardlaw, A. J. The role of air pollution in asthma. *Clin. Exp. Allergy 23, 81-96.*
7. Schwartz, J. Air pollution and hospital admissions for the elderly in Detroit, Michigan. *Am J Respir Crit Care Med 1994;150: 648-655.*
8. Daniel Slater, and Timothy V. Particulate air pollution and hospital emergency room visits for asthma in Seattle. *Am Rev Respir Dis 1993;147: 826-831.*
9. Eana H. Jaffe, Mendel E. Singer, and Alfred A. Rimm. Air pollution and emergency department visits for asthma among Ohio Medicaid recipients, 1991-1996. *Environmental Research 2003; 9:21-28*
10. Eli Ganor, Evsey Kosman, and Yoav Waiser. Emergency room visits of asthmatic children, relation to air pollution, weather, and airborne allergens. *Annals of Allergy, Asthma, & Immunology 1998; 81: 563-570.*
11. Burge H.A. Airborne-allergenic fungi. *Immunol. Allergy Clin. North Am 1989;9: 307-319.*

12. Agata, H., Yomo, A., and Orii, T. et al. Comparison of the MAST chemilluminescent assay system with RASAT and skin tests in allergic children. *Annals of Allergy 1993;70:153-157.*

13. Kam, K.-L. and Hsieh, K.-H et al. Comparison of three in vitro assays for serum IgE with skin testing in asthmatic children. *Annals of Allergy 1994; 73:329-336.*

14. Munro, D.A. Aeroallergens and childhood asthma in Melbourne. *Clean Air 10,42-45.*

15. Brown, H. M. and Jackson, F. Asthma and the weather. *Lancet 1983;21:630.*

16. Reid, M.J., Moss, R.B., and Nelson, B.L. Seasonal asthma in northern California: allergic causes and efficacy of immunotherapy. *Journal of allergy and clinical Immunology 1986;78: 590-600.*

17. Boulet, L.-P. Respiratory pathophysiologic responses. Influence of natural antigenic exposure on expiratory flows, methacholine responsiveness, and airway inflammation in mild allergic asthma. *Journal of Allergy and Clinical Immunology 1993;91:883-893.*

18. Rosas, I., and Mccartney, HA. Analysis of the relationships between environmental factors (aeroallergens, air pollution, and weather) and asthma emergency admissions to a hospital in Mexico City. *Allergy 1998;53:394-401.*

19. Mechelle B. Lierl, and Richard W. Hornung. Relationship of outdoor air quality to pediatric asthma exacerbations. *Annals of Allergy, Asthma, & Immunology 2003;90:28-33.*

20. McCullagh,P., and Nelder, J.A.Generalized Linear Models. *Champman and Hall, London, 1983.*

21. Hastie,T., and Tibshirani,R. *Generalized Additive Models. Chapman and Hall, London.1990*

22. Schwartz J. Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Canadian Journal of Statistics, 1994,4:471-87*

23. Adhikari, A., and Martuzevicius D. Performance of the Button Personal Inhalable Sampler for the measurement of outdoor aeroallergens. *Atmospheric Environment (submitted).*

24. Liang KY, and Zeger SL. *Longitudinal data analysis using generalized linear models. Biometrika 1986;73: 13-22.*

25. Carl Moones, Frank Harrell. Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis. *Springer New York, 2001.*

26. Salvaggio J, Hasselblad V. Seabury J, Heiderscheit LT. New Orleans asthma. II. Relationship of climatologic and seasonal factors to outbreaks. *J Allergy 1970; 45:257-65.*

27. Viswanath P. Kurup, Horng-Der Shen. Respiratory fungal allergy. *Microbes and Infection, 2000; 2: 1101-1110.*

28. Paul J. Beggs, Pollen and pollen antigen as triggers of asthma – what to measure? *Atmospheric Environment, 1998, Vol 32, 10: 1777-1783.*

29. The increasing trend of seasonal respiratory allergy in urban areas. G.D'Amato, G. Liccardi. *Allergy 2002; 57; Suppl. 71:35-36.*

30. Diaz-Sanchez D, Tsien A, Fleming J. Combined diesel exhaust particulate and ragweed allergen challenge markedly enhances human in vivo nasal ragweed-

specific IgE and skews cytokine production to a T helper cell 2-type pattern. *J Immunol 1997;158: 2406-2413.*

31. Takafuji S, Suzuki S. Enhancing effect of suspended particulate matter on the IgE antibody production in mice. *Int Arch Allergy Appl Immunol 1989; 90: 1-7.*

32. Takano H, Yoshikawa T. Eiesel exhaust particles enhance antigen-induced airway inflammation and local cytokine expression in mice. *Am J Respir Crit Care Med 1997; 156: 36-42.*

Appendix

1.Program to calculate the descriptive statistics for asthma hospital admission and aeroallergen concentration

```
libname lib "S:\common\weizhong\ms thesis\dataset";

data lib.admission;set admissiondata;
 if 365<=age_in_days <=6570;  *select patients with age from 1-18;
run;

proc sort data=lib.admission;by subject visit_start_date;
data admission;set lib.admission;by subject visit_start_date;
 if first.subject=0 and first.visit_start_date=0 then delete;
run;  *If a person had >1 asthma visits in a day, only 1 was counted;

proc sort data=admission;by visit_start_date;
proc means data=admission noprint;by visit_start_date;
 output out=lib.admmission n=admission;
run;                *daily admission number was calculated;

proc sort data=lib.admission;by date;
proc sort data=lib.pollen2002;by date;
data lib.pollenfinal;merge lib.pollen2002 lib.admission; by date;
 if pollen=. then delete;
*daily pollen data was combined with daily admission data;

proc sort data=lib.fungal2002;by date;
data lib.fungalfinal;merge lib.fungal2002 lib.admission; by date;
 if fungal=. then delete;
run;     *daily fungal data was combined with daily admission data;

proc sort data=lib.pollenfinal;by date;
proc sort data=lib.fungalfinal;by date;
data lib.aeroallergen; merge lib.pollenfinal lib.fungalfinal;by date;
run; *daily aeroallergen dataset was created;
```

```
data start;
 input begin mmddyy10.;
 format begin mmddyy10.;
cards;
03/05/2002
;
data lib.aeroallergen;  *create month and time variable;
 if _n_=1 then set start;
 set lib.aeroallergen;
 month=month(date);
 time=date-begin+1;
 drop begin;
run;


proc sort data=lib.aeroallergen;by  month;        *create summary
results;
proc means data=lib.aeroallergen noprint;by month;
 var admission temperature humidity pollen grass oak ragweed fungal alt
     cla asp;
 output out=summary mean=meanadmission meantemp meanhumid meanpollen
        meangrass meanoak meanragweed meanfungal meanalt meancla meanasp
        std=stdadmission stdtemp stdhumid stdpollen stdgrass stdoak
        stdragweed stdfungal stdalt stdcla stdasp;
run;

proc corr data=lib.aeroallergen;
 var temperature humidity pollen grass ragweed oak;
                                  *correlation coefficients were
calculated;
proc corr data=lib.aeroallergen;
 var temperature humidity fungal alt cla asp;
run;
```

2. Program to model seasonal, long-term temporal patterns and weather dependence

```
%macro one(a=);
proc genmod data=lib.aeroallergen;
 class month weekday;
 model admission=&a /dist=poisson type1 type3 pscale;
 run;
%mend;
%one(a=time);
%one(a=temperature);
%one(a=humidity);
%one(a=weekday);
%one(a=month);
run;

*linear and quadratic time term were added to control temporal pattern;

proc genmod data=lib.aeroallergen;
 model admission=time time*time/dist=poisson type1 type3 pscale;
run;*deviance=285.2 with df=151;

*restricted cubic spline funtion were used to capture secular trend;

proc univariate data=lib.aeroallergen;var time;
 output out=time pctlpts=0 5 27.5 50 72.5 95 pctlpre=t;run;
data lib.aeroallergen;set lib.aeroallergen;
 t1=10;t2=64;t3=126.5;t4=178;t5=232;
 x1=time;
 x2=max(0,(x1-t1))**3-max(0,(x1-t4))**3*(t5-t1)/(t5-t4)+max(0,(x1-
    t5))**3*(t4-t1)/(t5-t4);
 x3=max(0,(x1-t2))**3-max(0,(x1-t4))**3*(t5-t2)/(t5-t4)+max(0,(x1-
    t5))**3*(t4-t2)/(t5-t4);
 x4=max(0,(x1-t3))**3-max(0,(x1-t4))**3*(t5-t3)/(t5-t4)+max(0,(x1-
    t5))**3*(t4-t3)/(t5-t4);

 proc genmod data=lib.aeroallergen;
  model admission=x1 x2 x3 x4/dist=poisson type1 type3 pscale;
 run;*deviance=234.4, df=149, thus cubic spline was chosen;
```

```
*x1 x2 x3 x4 month were added first, since they were used to control
the seasonal  and temporal pattern, then temperature, humidity and
weekday effects were added;

*macro language were used to divide temperature and humidity into
different ranges from 5 to 9;
 options macrogen;
 %macro category (weather=,n=,data=,climate=,a=);
  %let percents=;
 data _null_;
  do i=0 to &n;
   call symput('percents',symget('percents')||'
'||trim(left(round(i*100/&n,.01))));
   end;
proc univariate data=lib.aeroallergen noprint;
 var &weather;
 output  out=stats pctlpre=p_ pctlpts=&percents;
proc transpose data=stats out=temporary;
 %let names=;
data _null_; set temporary;
 call symput('names',symget('names')||' '||trim(_name_));
 %let categor=;
 data temporary1;
  if _n_=1 then set stats;
set lib.aeroallergen;
 &weather&a=(%scan(&names,1,%str( ))<=&weather<=%scan(&names,2,%str(
)));
 %do i=2 %to &n;
 &weather&i=(%scan(&names,&i,%str( ))<&weather<=%scan(&names,&i+1,%str(
)));
 %let categor=&categor cat&i;
 %end;
data &data;set temporary1;
 &climate=0;
 %do i=1 %to &n;
 &climate=&climate+&i*&weather&i;
 %end;
%mend;
```

```
%category(weather=temperature,n=5,data=temp5,climate=temp,a=1);
%category(weather=humidity,n=5,data=humid5,climate=humid,a=1);
%category(weather=temperature,n=6,data=temp6,climate=temp,a=1);
%category(weather=temperature,n=7,data=temp7,climate=temp,a=1);
%category(weather=temperature,n=8,data=temp8,climate=temp,a=1);
%category(weather=temperature,n=9,data=temp9,climate=temp,a=1);
%category(weather=humidity,n=6,data=humid6,climate=humid,a=1);
%category(weather=humidity,n=7,data=humid7,climate=humid,a=1);
%category(weather=humidity,n=8,data=humid8,climate=humid,a=1);
%category(weather=humidity,n=9,data=humid9,climate=humid,a=1);
run;

*marco to clean the data;
options macrogen;
%macro cleaning(n=);
data cleaning;
 %do i=5 %to &n;
data temp&i;set temp&i;
    *  keep  date  weekday  month  admission  time  temp  x1  x2  x3  x4
temperature humidity;
data humid&i;set humid&i;
keep date humid;
%end;
%mend;
%cleaning (n=9);
run;

*macro to merge the datasets;
%macro merge (n=);
data a;
 %do i=5 %to &n;
 proc sort data=temp&i;by date;
 proc sort data=humid&i;by date;
 data category&i;merge temp&i humid&i;by date;
%end;
%mend;
```

```
%merge (n=9);
run;

* use macro to select which category division was most appropriate;
ods trace on;
options macrogen;
%macro group(n=);
data a;
 %do i=5 %to &n;
proc genmod data=category&i;
 class temp;
 model admission= temp /dist=poisson type1 type3 pscale;
 ods output modelfit=fit&i;
 ods output type1=type1group&i;
 ods output type3=type3group&i;
 %end;
 %mend
 %group(n=9);
 run; ;*for temperature, category 6 was selected, because it improved
model fit most significantly;*

ods trace on;
options macrogen;
%macro group(n=);
data a;
 %do i=5 %to &n;
proc genmod data=category&i;
 class humid;
 model admission=humid /dist=poisson type1 type3 pscale;
 ods output modelfit=fit&i;
 ods output type1=type1group&i;
 ods output type3=type3group&i;
 %end;
 %mend;
 %group (n=9);
 run; *humid is not significant in model fit improvement and thus was
       excluded;
```

```
data lib.category6;set category6; *6 dummies for temperature were
added;
keep date temp;
proc sort data=lib.category6;by date;
proc sort data=lib.aeroallergen;by date;
data lib.aeroallergenfinal;merge lib.aeroallergen lib.category6;by
date;
run;*lib.aeroallergenfinal would be used in the subsequent analysis;

*Model was checked to see if temporal pattern and weather dependence
has been controlled;
proc genmod data=lib.aeroallergenfinal;
 class week month weekday temp;
 model admission=x1 x2 x3 x4 month temp weekday/dist=poisson type1
  type3   pscale obstats;
 ods output obstats=modelcheck; *Diagnostic plots would be generated;
run;
data temperature;set lib.aeroallergen;
 keep temperature time;
data lib.modelcheck; merge modelcheck temperature;
run; *lib.modelcheck would be exported to S-plus to generate diagnostic
      plots; *plot was ok;



*Ragweed and Oak/Cedar pollens have their unique  seasonal appearance,
different models were fitted;

data ragweed;set lib.aeroallergen;
 if month in (8,9,10);
 run;

data start;
 input begin mmddyy10.;
 format begin mmddyy10.;
cards;
```

```
08/01/2002
;


data lib.ragweed; *set time variable;
 if _n_=1 then set start;
 set ragweed;
 time=date-begin+1;
 keep date ragweed ragweed1 ragweed2 ragweed3 ragweed4 ragweed5
      temperature  humidity month admission time week weekday;
run;

data oak;set lib.aeroallergen;
 if month in(3,4,5,6);
run;

data start;
 input begin mmddyy10.;
 format begin mmddyy10.;
cards;
03/05/2002
;
data lib.oak;*set time variable;
 if _n_=1 then set start;
 set oak;
 time=date-begin+1;
 keep date oak oak1 oak2 oak3 oak4 oak5 temperature humidity month
      admission time week weekday;
run;

* knots selection for Ragweed pollen;
*with small sample size, 3 knots for restricted cubic spline was
enough;
proc univariate data=lib.ragweed;var time;
 output out=time pctlpts=10 50 90pctlpre=t;

data lib.ragweed;set lib.ragweed;
 t1=13.5;t2=48.5;t3=84.5;
```

```
 x1=time;
 x2=max(0,(x1-t1))**3-max(0,(x1-t2))**3*(t3-t1)/(t3-t2)+max(0,(x1-
  t3))**3*(t2-t1)/(t3-t2);
run;

*temperature was divided into 3 categories after the same selection
procedure was applied;


   *model was checked see if temporal pattern and weather dependence
has been controlled;
proc genmod data=lib.ragweedfinal;
 class month weekday temp;
 model admission=x1 x2  month temp weekday/dist=poisson type1 type3
  pscale obstats;
 ods output obstats=modelcheckragweed;
run; *plot looks ok;


*knots selection for Oak/Cedar Pollen;

proc univariate data=lib.oak;var time;
 output out=time pctlpts=10 50 90 pctlpre=t;
run;
data lib.oak;set lib.oak;
 t1=10;t2=56;t3=100;
 x1=time;
 x2=max(0,(x1-t1))**3-max(0,(x1-t2))**3*(t3-t1)/(t3-t2)+max(0,(x1-
  t3))**3*(t2-t1)/(t3-t2);
run;

*temperature was divided into 6 categories after the same selection
procedure was applied;

*model was checked to see if temporal pattern and weather dependence
has been controlled;
proc genmod data=lib.oakfinal;
 class week month weekday temp;
```

```
 model admission=x1 x2  month temp weekday/dist=poisson type1 type3
  pscale  obstats;
 ods output obstats=modelcheckoak;
run; *plot looks ok;
```

3. Genmod Program

```
%macro genmod(allergen=,residual=); *aeroallergens entered model
singly;
 proc genmod data=lib.aeroallergenfinal;
  class month weekday temp ;
  model    admission=x1    x2    x3    x4    month        temp    weekday
  &allergen/dist=poisson  type1 type3 pscale obstats;
  ods output obstats=&residual;
 proc univariate data=&residual normal; var resdev;
 proc autoreg data=&residual; *residual serial correlation was checked;
  model resdev=/nlag=2;
%mend;
%genmod(allergen=pollen ,residual=pollen);
%genmod(allergen=pollen1,residual=pollen1);
%genmod( allergen=pollen2,residual=pollen2);
%genmod(allergen=pollen3,residual=pollen3);
%genmod(allergen=pollen4,residual=pollen4);
%genmod(allergen=pollen5,residual=pollen5);
%genmod(allergen=grass,residual=grass);
%genmod(allergen=grass1,residual=grass1);
%genmod(allergen=grass2,residual=grass2);
%genmod(allergen=grass3,residual=grass3);
%genmod(allergen=grass4,residual=grass4);
%genmod(allergen=grass5,residual=grass5);
%genmod(allergen=fungal,residual=fungal);
%genmod(allergen=fungal1,residual=fungal1);
%genmod(allergen=fungal2,residual=fungal2);
%genmod(allergen=fungal3,residual=fungal3);
%genmod(allergen=fungal4,residual=fungal4);
%genmod(allergen=fungal5,residual=fungal5);
%genmod(allergen=alt,residual=alt);
%genmod(allergen=alt1,residual=alt1);
%genmod(allergen=alt2,residual=alt2);
%genmod(allergen=alt3,residual=alt3);
%genmod(allergen=alt4,residual=alt4);
%genmod(allergen=alt5,residual=alt5);
%genmod(allergen=asp,residual=asp);
```

```
%genmod(allergen=asp1,residual=asp1);
%genmod(allergen=asp2,residual=asp1);
%genmod(allergen=asp3,residual=asp3);
%genmod(allergen=asp4,residual=asp4);
%genmod(allergen=asp5,residual=asp5);
%genmod(allergen=cla,residual=cla);
%genmod(allergen=cla1,residual=cla1);
%genmod(allergen=cla2,residual=cla2);
%genmod(allergen=cla3,residual=cla3);
%genmod(allergen=cla4,residual=cla4);
%genmod(allergen=cla5,residual=cla5);
run;

*Ragweed and Oak/Cedar pollens have their unique seasonal appearance,
different models were fitted;

*Ragweed pollen with different lag days were entered singly into the
model;
%macro ragweed(allergen=,residual=);
proc genmod data=lib.ragweedfinal;
 class month weekday  temp;
 model admission=x1 x2 month temp weekday &allergen/dist=poisson type1
  type3
 pscale obstats;
 ods output obstats=&residual;
 proc univariate data=&residual normal; var resdev;
 proc autoreg data=&residual; *residual serial correlation was checked;
  model resdev=/nlag=2;
%mend;
%ragweed(allergen=ragweed,residual=ragweed);
%ragweed(allergen=ragweed1,residual=ragweed1);
%ragweed(allergen=ragweed2,residual=ragweed2);
%ragweed(allergen=ragweed3,residual=ragweed3);
%ragweed(allergen=ragweed4,residual=ragweed4);
%ragweed(allergen=ragweed5,residual=ragweed5);
run;
```

```
*Oak/Cedar pollen with different lag days were entered singly into the
model;
%macro oak(allergen=,residual=);
proc genmod data=lib.oakfinal;
 class month weekday  temp;
 model admission=x1 x2 month temp weekday &allergen/dist=poisson type1
  type3 pscale obstats;
  ods output obstats=&residual;
 proc univariate data=&residual normal; var resdev;
 proc autoreg data=&residual; *residual serial correlation was checked;
  model resdev=/nlag=2;
%mend;
%oak(allergen=oak,residual=oak);
%oak(allergen=oak1,residual=oak1);
%oak(allergen=oak2,residual=oak2);
%oak(allergen=oak3,residual=oak3);
%oak(allergen=oak4,residual=oak4);
%oak(allergen=oak5,residual=oak5);
run;
```

4.Gam Program

```
*S-Plus was used to search for the appropriate smoothing parameter for
time and temperature, df=9 and df=6 were chosen respectively for time
and temperature;

options macrogen;
%macro gam(allergen=);
proc gam data=lib.aeroallergenfinal;
 class weekday;
 model admission= spline(time,df=9) spline(temperature,df=6)
param(weekday  &allergen)/dist=poisson;
%mend;
%gam(allergen=pollen);
%gam(allergen=pollen1);
%gam(allergen=pollen2);
%gam(allergen=pollen3);
%gam(allergen=pollen4);
%gam(allergen=pollen5);
%gam(allergen=grass);
%gam(allergen=grass1);
%gam(allergen=grass2);
%gam(allergen=grass3);
%gam(allergen=grass4);
%gam(allergen=grass5);
%gam(allergen=fungal);
%gam(allergen=fungal1);
%gam(allergen=fungal2);
%gam(allergen=fungal3);
%gam(allergen=fungal4);
%gam(allergen=fungal5);
%gam(allergen=alt);
%gam(allergen=alt1);
%gam(allergen=alt2);
%gam(allergen=alt3);
%gam(allergen=alt4);
%gam(allergen=alt5);
%gam(allergen=asp1);
```

```
%gam(allergen=asp2);
%gam(allergen=asp3);
%gam(allergen=asp4);
%gam(allergen=asp5);
%gam(allergen=cla);
%gam(allergen=cla1);
%gam(allergen=cla2);
%gam(allergen=cla3);
%gam(allergen=cla4);
%gam(allergen=cla5);
run;

*Ragweed and Oak/Cedar pollens have their unique seasonal appearance,
different models were fitted;

*gam model for Ragweed pollen;
*df=6 and df=3 were chosen respectively for time and temperature for
their smoothing splines;
%macro gamragweed(allergen=);
proc gam data=lib.ragweedfinal;
 class weekday;
 model admission= spline(time,df=6) spline(temperature,df=3)
  param(weekday &allergen)/dist=poisson;
%mend;
%gamragweed(allergen=ragweed);
%gamragweed(allergen=ragweed1);
%gamragweed(allergen=ragweed2);
%gamragweed(allergen=ragweed3);
%gamragweed(allergen=ragweed4);
%gamragweed(allergen=ragweed5);
run;

*gam model for Oak/Cedar Pollen;
*df=6 was chosen for smoothing splines for both time and temperature ;
%macro gamoak(allergen=);
proc gam data=lib.oakfinal;
 class weekday;
```

```
 model admission= spline(time,df=6)  spline(temperature,df=6)
  param(weekday  &allergen)/dist=poisson;
%mend;
%gamoak(allergen=oak);
%gamoak(allergen=oak1);
%gamoak(allergen=oak2);
%gamoak(allergen=oak3);
%gamoak(allergen=oak4);
%gamoak(allergen=oak5);
run;
```