Comparing Exposure Metrics for the Effects of Fine Particulate Matter on Emergency Hospital Admissions

Elizabeth Mannshardt[1], Katarina Sucic[1], Wan Jiao[1],
Francesca Dominici[2], H. Christopher Frey[1], Brian Reich[1], and Montserrat Fuentes[1]

[1]North Carolina State University
[2]Harvard University

Oct 19, 2012

Abstract

A crucial step in an epidemiological study of the effects of air pollution is to accurately quantify exposure of the population. In this paper, we investigate the sensitivity of the health effects estimates associated with short-term exposure to fine particulate matter with respect to three potential metrics for daily exposure: ambient monitor data, estimated values from a deterministic atmospheric chemistry model, and stochastic daily average human exposure simulation output. Each of these metrics has strengths and weaknesses when estimating the association between daily changes in ambient exposure to fine particulate matter and daily emergency hospital admissions. Monitor data is readily available, but is incomplete over space and time. The atmospheric chemistry model output is spatially and temporally complete, but may be less accurate than monitor data. The stochastic human exposure estimates account for human activity patterns and variability in pollutant concentration across microenvironments, but requires extensive input information and computation time. To compare these metrics, we consider a case study of the association between fine particulate matter and emergency hospital admissions for respiratory cases for the Medicare population across three counties in New York. Of particular interest is to quantify the impact and/or benefit to using the stochastic human exposure output to measure health exposure to fine particulate matter. Results indicate that the stochastic human exposure simulation output indicates approximately the same increase in relative risk associated with emergency admissions as using a chemistry model or monitoring data as exposure metrics. However, the stochastic human exposure simulation output and the atmospheric chemistry model both bring additional information which helps to reduce the uncertainly in our estimated risk.

Keywords: Health effects, air polluation, particulate matter, ambient monitoring data, CMAQ, exposure models, SHEDS, $PM_{2.5}$

1 Introduction

Numerous studies have shown the positive association between particulate matter and adverse human health effects - Dominici, Peng and Bell [1]; Pope et al [2], Bell et al [3], and Ostro et al [4] for respiratory effects, among others. In other examples, air pollutants are linked to a significant increase in respiratory deaths in Ostro et al [5] and Braga et al [6]. Holloman et al [7] relates $PM_{2.5}$ exposure to cardiovascular mortality, Braga et al [6] and Hoek et al [8] relate pollution exposure to cardiovascular disease, and Pope et al [9], Dockery et al [10], Dockery and Pope, [11], and Fuentes et al [12] related pollutant exposure to mortality and morbidity in general. The Environmental Protection Agency is "concerned about particles that are 10 micrometers in diameter or smaller because those are the particles that generally pass through the throat and nose and enter the lungs" [13]. *Fine particles*, $PM_{2.5}$, such as those found in smoke and haze, are defined as 2.5 micrometers in diameter and smaller. Once inhaled, "particle pollution - especially fine particles - contain microscopic solids or liquid droplets that are so small that they can get deep into the lungs and cause serious health problems" [13]. According to the EPA, particle pollution exposure has been specifically linked to a variety of problems including premature death in people with heart of lung disease, nonfatal heart attacks, irregular heartbeat, aggravated asthma, decreased lung function, and increased respiratory symptoms, such as irritation of the airways, coughing or difficulty breathing [13]. The EPA warns that "people with heart or lung diseases, children, and older adults are the most likely to be affected by particle pollution exposure" [13].

The aim of this paper is to estimate the short-term health effects of ambient exposure to fine particulate matter on population health outcomes, where $PM_{2.5}$ levels are obtained from three different exposure metrics. The National Morbidity, Mortality, and Air Pollution Study (NMMAPS) [14] describes statistical methods for estimating the risk of various health effects in the Medicare population due to fine particulate matter exposure. A challenging aspect of air pollution studies is properly quantifying the exposures of individuals in the population. We investigate the sensitivity of the estimated health effect of three potential exposure metrics for fine particulate matter: ambient monitor data (AQS), estimated air quality concentrations from a deterministic atmospheric chemistry model - the Community Multi-scale Air Quality modeling system (CMAQ) described in Byun and Schere [15], and simulated individual daily average exposure based on the Stochastic Human Exposure and Dose Simulation model (SHEDS-PM). Each metric considered here has different strengths and weaknesses. AQS monitoring data is readily available but is spatially and temporarily incomplete, whereas CMAQ output is spatially and temporally complete, but may be less accurate. SHEDS-PM output accounts for population exposure variability but requires extensive input and is computationally expensive. CMAQ serves as a surrogate for directly measuring ambient pollution exposure and SHEDS-PM is a surrogate for population exposoure to fine particulate matter. The outcome considered here is emergency respiratory disease hospital admissions for Medicare patients age 65 and older for the period 2002-2006.

SHEDS-PM can provide information about short-term population ambient exposure.

Several recent papers have used Bayesian hierarchical models to incorporate output from an exposure simulator as predictors of various health responses (Calder et al [16], Berrocal et al [17], Blangiardo et al [18], and Reich et al [19]). This paper focuses in particular on the comparison of health effects models incorporating SHEDS-PM output as explanatory variables, as it is of interest to the scientific community to understand the possible benefits to be gained from population exposure information.

A limitation of many studies of adverse human health effects is that a single exposure value is used for all individuals whereas personal exposure can vary greatly. While direct measurements of individual exposure are not available with sufficient spatial and temporal coverage to enable comparison with health effects data at the scale evaluated here, SHEDS-PM estimates population distributions of inter-individual variability in daily average exposure using information about human activity patterns and living environments, as well as census data. In this paper, we present a comparison of exposure metrics utilizing a modeling framework to capture the population exposure information. The comparison is showcased in a simulation study as well as an application on emergency hospital room admissions for respiratory disease.

First, Section 2 describes the exposure metrics and the data used for the application. Then Section 3 details the methodology and models, and Section 4 outlines the simulation study. Section 5 describes the application and Section 6 explains the model results for the effect of $PM_{2.5}$ on emergency hospital admissions. Finally, a discussion comparing the exposure metrics is presented in Section 7.

2 Data

In this application we focus on three counties in the New York City area: Bronx, Queens, and New York Counties, for the years 2002-2006. The outcome of interest is emergency respiratory disease hospital admissions for Medicare patients age 65 and older, obtained from emergency hospital admissions data created from the Medicare Part A and Medicare Denominator files, where daily time series of hospitalizations were constructed for each county as described in Peng et al [20]. Respiratory admissions were classified based on "*ICD-9* codes including chronic obstructive pulmonary disease (490-448 and respiratory tract infections (464-466, 480-497)" [20]. For each outcome, only the primary diagnosis for the hopital admission was considered as the basis for inclusion and daily time series of hospitalization rates were constructed by cause for each county by summing the number of emergency hospital admissions for each day in each county [20].

AQS monitor data measurements are observed approximately every third day, and the resulting data product is an aggregate county measurement averaged over the stations located in each county. Another important source of $PM_{2.5}$ over large areas can be obtained from the three-dimensional (3-D) regional scale air quality models such as the U.S. EPA Community Multiscale Air Quality (CMAQ) modeling system (Binkowski and Rosell [21]; Byun and Schere [15]). CMAQ output is spatially and temporally complete,

but may be less accurate than monitor data as it provides estimates of the pollutants on a grid. CMAQ is a deterministic chemistry model based on stochastic differential equations that describe that underlying chemistry [22]. CMAQ simulations over an airshed of interest provide gridded hourly concentrations and dry/wet deposition fluxes of major air pollutants such as ozone and fine particles at a 12×12 km^2 resolution for the entire Eastern United States from 2002-2006. CMAQ has various sources of uncertainties, including the support space. CMAQ was designed to have multi-scale capabilities for both regional and urban scale modeling, however many of the hydrostatic static assumptions valid on a regional scale cannot be made on an urban scale. Non-hydrostatic equations are more appropriate for finer urban scales, though the target grid resolutions and domain sizes for CMAQ range spatially and temporally over several orders of magnitude. [23] AQS monitoring data is compared to a CMAQ data product in Chang et al [24] to relate fine particulate matter to pre-term birth, and Bravo et al [25] consider CMAQ as a metric for pollutant exposure in epidemiological studies.

SHEDS-PM is a population exposure model for particulate matter developed by the US Environmental Protection Agency. SHEDS-PM employs a probabilistic approach to estimate distributions of inter-individual variability in outdoor and indoor microenvironmental $PM_{2.5}$ exposures for a simulated population based on ambient air quality and human activity data (Burke) [26], such as workplace or residential environment and exposure through cooking and smoking. The human activity data are based on the Consolidated Human Activity Database (CHAD) [27], which is based on over $22,000$ daily dairies of participants documenting time spent in various micro-environments. Figure (8) provides a schematic of this algorithm. A Monte Carlo sampling scheme is used to estimate $PM_{2.5}$ concentrations and their uncertainties in each microenvironment. For example, for the residential microenvironment, a single-compartment, steady state mass balance equation is used. CMAQ output is used as input for the ambient concentration. Key factors that influence the fraction of ambient $PM_{2.5}$ concentration which penetrates and remains in the residential microenvironment are: (1) air exchange rate (ACH) [28]; (2) penetration factor [29]; and (3) deposition rate [29].

Lognormal ACH distributions were developed based on data for New York City from the Toxic Exposure Assessment: A Columbia-Harvard Study (TEACH-NYC) for cold and warm day temperature categories (Jones et al [30]). Days with daily average temperatures less than 65°F are defined as "cold," whereas days >= 65°F. are defined as "warm." The cold and warm ACH distributions were applied to each simulated day depending on the daily average temperature. Inter-individual geographic variability in exposures is described in Cao and Frey [31] and [29]. The output includes predicted daily average values of ambient (E_a), non-ambient (E_{na}), and total exposure (E_t) for each simulated individual for each simulated day, and time spent in each microenvironment. An aggregated county sample mean and standard deviation were calculated for each county. Ratios of E_a/C for each simulated individual were calculated from daily average ambient exposure divided by input ambient $PM_{2.5}$ concentration (C).

2.1 Data Processing

Daily hospital admissions data are available on a county level, thus it was necessary to convert AQS, CMAQ, and SHEDS-PM to aggregates on a county level. CMAQ output is on a 12 $\times$ 12 kilometer grid. To aggregate to the county level, a weighted average across the grids was calculated based on the proportion of each county on each grid. We use a database of fine $PM_{2.5}$ and ozone monitoring data from AQS, modeled CMAQ output from the EPA, and health data from Medicare billing claims (as detailed in Peng et al [20] and Dominici [1]), as well as daily weather conditions. We also have SHEDS-PM daily total particulate matter exposure simulated for approximately 50,000[1] people age 65 and older for this same time-frame. To aggregate to the county level, the tracts within each county were averaged. Our response is daily respiratory (RESP) disease emergency hospital admissions for Medicare patients for the period 2002-2006.

We standardized daily fine $PM_{2.5}$ by subtracting the sample mean and dividing by the sample standard deviation across time for each of the three counties. Additionally, we tested for outliers by isolating days that were six times the interquartile range above the median values. However, the five days that met this criteria were kept because they were deemed reasonable given the pollutant and time of year that they occurred. Lag terms for ozone and particulate matter were created using the one day lag for ozone and $PM_{2.5}$. Total particulate matter exposure includes particles of ambient and non-ambient origin, taking into account air exchange rate, penetration, deposition, smoking status and cooking habits. We use only ambient exposure for comparison with AQS and CMAQ. The average and the variance of $PM_{2.5}$ exposures were calculated from the SHEDS-PM simulation for inclusion in the individual exposure model in Equation (2).

There were some instances of missing data. For New York County, where ozone readings were missing, the missing ozone values were infilled with CMAQ ozone output. In the CMAQ output, there were five non-consecutive days where fine particulate matter concentration was missing. These missing values were replaced with the average values from the dates preceding and immediately following the missing dates. The infilled values represent less than one third of one percent of the overall CMAQ particulate matter output.

3 Methodology

Many studies (two papers by Dominici, [32] and [33], and Peng et al [34]) have illustrated the potential confounders associated with air pollution and health effects, and the importance of adjusting for these effects. We employ the semi-parametric method outlined in Peng, Dominici, and Louis [34] to adjust for seasonal and long-term trends by incorporating natural splines. The use of nonparametric smoothing for health effects

[1]The simulation was run to collect information for 8.3% of the population using census values. This corresponds to approximately 50,000 people each year, spread proportionally over the counties, above the age of 65.

time series models was introduced in Schwartz [35], where smooth functions were used for time, temperature, dewpoint, and PM_{10}. The smooth function of time accounts for potential confounding factors which vary smoothly over time. Natural splines are utilized to control for long-term trends and seasonality over time. Weather variables such as temperature and relative humidity are also considered confounders. Therefore splines in temperature and relative humidity are incorporated into the model, as well as linear and quadratic terms in time. Also of interest are possible confounding long-term trends due to delayed onset of hospital admissions after exposure. Thus a confounding term is included for the 1-day lag for ozone, as well as temperature where the mean value is taken over the preceding 3-day period.

3.1 Models

Define Y_t as the total number of events, i.e. the number of emergency respiratory admissions, on day t, across all three counties. As potential confounders, we use a linear and quadratic fit in time, and spline fits in maximum daily temperature (temp_t) and average daily relative humidity (hum_t). Additional non-pollutant confounders considered are the temperature lag defined as the average temperature over the previous three days ($\overline{\text{mean(temp)}}_t$), ozone, and day of the week (dow), where dow has six levels corresponding to the calendar days of the week, with Saturday as the baseline exposure. There are certainly other covariates and confounders that could be considered for the modeling of human health effects such as emergency hospital admissions in the presence of $PM_{2.5}$. The focus of this paper is the comparison of the three different available exposure metrics - measured air quality (AQS), modeled air quality (CMAQ), and modeled individual exposure(SHEDS-PM). Thus the focus here is to create a base model that captures the basic characteristics of a pollutant model that enables the comparison of these metrics and their effectiveness at providing exposure information for $PM_{2.5}$. In the application presented here, $PM_{2.5}$ is standardized by subtracting the mean and dividing by the standard deviation for each of the three counties.

We include the ambient concentrations of fine particulate matter from either AQS monitor data or from CMAQ output. In models with the AQS monitoring data and the CMAQ output, the exposures are ambient concentrations and are considered constant for the entire population and are denoted PM_t for day t. The counts are modeled as Poisson with

$$\begin{aligned} \log[E(Y_t)] &= \log N_t + \beta_0 + s(\text{temp}_t; d_1) + s(\overline{\text{hum}}_t; d_2) + \beta_1 \text{t} + \beta_2 \text{t}^2 \\ &+ \beta_3 \overline{\text{mean(temp)}}_t + \beta_4 \text{ozone}_t + \beta_{dow} \text{dow}_t \\ &+ PM_t \beta_{PM} \end{aligned} \tag{1}$$

where $s(;d)$ is the natural spline basis expansion with d degrees of freedom, chosen as explained in Section 6.3. $E(Y_t)$ represents the expected number of counts for time t, and $\beta_{dow}\text{dow}_t$ is a vector $(\beta_S \mathbf{I}(\text{Sun}) + ... + \beta_F \mathbf{I}(\text{Fri}))$ corresponding to the calendar days of the week, where Saturday is considered the baseline level for pollutant exposure. This standard model assumes that there are no interactions between covariates, and includes

an offset term for Poisson models, $\log N_t$, where N_t is the number of Medicare participants in the study.

The analysis incorporating estimated personal exposure is approached differently, as the SHEDS-PM personal exposure model allows us to consider exposure at an individual level. SHEDS-PM does not assume that the exposure is the same for all individuals in the population, and the health model must be modified accordingly. If the exposure distribution on day t is estimated by SHEDS-PM to have mean m_t and variance v_t, then the expected number of counts can be modeled as in Reich, Fuentes, and Burke (2009) [19]. Note that if the variance of the exposure distribution is zero, then this reduces to the ambient concentration model with exposure $PM_t = m_t$. Here we are considering the lag-term for the mean personal exposure, m_{t-1}, as indicated by Braga 2001 [6].

$$\begin{aligned} \log[E(Y_t)] &= \log N_t + \beta_0 + s(\text{temp}_t; d_1) + s(\overline{\text{hum}}_t; d_2) + \beta_1 \text{t} + \beta_2 \text{t}^2 \\ &+ \beta_3 \overline{\text{mean(temp)}}_t + \beta_4 \text{ozone}_t + \beta_{dow} \text{dow}_t \\ &+ \alpha_{PM} m_{t-1} + \frac{1}{2} \alpha_{PM}^2 v_{t-1} \end{aligned} \tag{2}$$

The final term $\alpha^2 v_t$ accounts for variation in exposure across the population. The derivation of the modeling formulation is described in Reich et al 2009 [19]. An offset term, $\log \text{N}_T$, for Poisson count models is also included.

We carry out the analysis using Bayesian methods. The advantages of a Bayesian framework in pollutant effect models has been shown in multiple studies as well as utilized in the studies referenced in Section 1. Dominici (2002) [33] outlined the advantages to a Bayesian approach in modeling air pollution, Choi et al (2009) [36] uses a Bayesian framework to model $\text{PM}_{2.5}$ over space and time, Blangiardo et al (2009) [18] implement a Bayesian framework to relate individual level data from activity diaries to particulate matter exposure, and Reich et al (2009) [19] relates fine particulate matter, $\text{PM}_{2.5}$, to mortality using the SHEDS-PM simulated exposure. Reich et al introduce a Baysian model that incorporates the exposure distributions to account for variability in exposure across the population, which is the methodology considered here.

A Bayesian analysis begins by specifying a prior distribution for each model parameter which quantifies the information about parameter before observing the data. After observing the data, we have two sources of information, the data's likelihood and the prior, which are combined using Bayes' theorem to give the posterior distribution [37]. The posterior distribution represents the current state of knowledge based on all available information and is used for inference. A crucial step in a Bayesian analysis is selecting appropriate priors for model parameters. We use normal priors with mean zero and large variance for the coefficient parameters to allow for a non-informative prior. Markov Chain Monte Carlo methods are used to sample from the conditional distribution. A burn-in of 5,000 is discarded and 20,000 posterior draws are obtained for inference. Convergence was confirmed using trace-plots.

4 Simulation

A simulation study is conducted to test the power of detecting a relative risk signal from the three exposure metrics defined above. A health outcome data set, Z, of simulated health data is generated using random draws from a Poisson distribution with a linear mean function in the confounders, simulated values for the daily mean exposure M_t, and specified values for the variance V of the daily individual exposures. The expected number of simulated hospital admissions on day t can be expressed through the log relationship:

$$\log[E(Z_t)] = \log N_t + \beta_0 + \beta_1 \text{dew}_t + \beta_2 \text{temp}_t + \beta_3 \text{dow}_t + \beta_4 \text{ozone}_t \tag{3}$$

where a basic structure for the confounders considered was fit with $\beta = (0.1, 0.2, 0.3, 0.4)$ where dew_t = dewpoint and dow_t = day of week is an indicator for weekday, weekend, or holiday. An offset term for Poisson count models is also included. Z_t is simulated using the R function `rpois` as

$$Z_t = \texttt{rpois}(n, \exp(X\beta + M_t\alpha + \frac{1}{2}\alpha^2 V) \tag{4}$$

The mean M_t of the daily individual exposures is simulated according to the relationship:

$$M_t = C_t \cdot \exp(R) \tag{5}$$

with input C_t, of the observed 1-day lag AQS ambient exposure for day t and R representing random draws from a normal distribution with mean $E_{mn} = 0$ and standard deviation $E_{sd} = 0.88$. A constant variance V is used for purposes of simplicity. $V = 0.3$ is chosen as the average of the estimated personal exposure variances v_ϵ, and a larger variance, $V = 1$, was tested as well for a robustness comparison. $N_1 = 1623$ simulated hospital admission counts were generated, utilizing the 1623 available AQS observations over the 5-year time period.

In Equation (5) the observed AQS observations, denoted by C above, are used as to generate simulated personal exposure distributions. The data, Z_{N_1}, is generated to have correlation $\text{corr}(M_t, C_t) = 0.7$ where M_t is the mean exposure and C_t is the input AQS on day t, which is consistent with the correlation of AQS and SHEDS-PM as observed in New York counties data. A correlation of $r = 0.7$ corresponds to an approximate standard deviation of $E_{sd} = 0.88$ for the mean exposure. For each dataset we test the null hypothesis that the $PM_{2.5}$ effect on the relative risk is zero - i.e., α is not significantly different from 0. Other possible values of E_{sd} are also considered, as well as additional values of α and V, for robustness. The power of detecting the individual effect α with the distributional component $\frac{1}{2}\alpha^2 V$ described in Equation 3 is compared to the power of detecting the effect β_{PM} of fine particulate matter according to the model described in Equation (1), with the reduced set of confounders for simplicity. Table 1 displays the empirical power over $1,000$ simulations across a reasonable spectrum of possible values of α and standard deviation, E_{sd}. For each simulation 5,000 posterior samples were drawn after a burn-in of 500 using non-informative normal priors with mean 0 and

a large variance of 100 with convergence diagnostics checked via a sampling of trace-plots.

Table 1 and Figure 8 show that as the strength of the effect for $PM_{2.5}$ increases, the model incorporating the individual exposure has greater power than the model utilizing the ambient AQS data. The difference in power is significant for the standard deviation $E_{sd} = 0.88$ for the mean exposures, which is the most realistic scenario as a standard deviation of $E_{sd} = 0.88$ captures the observed dependence between AQS and SHEDS-PM. The personal exposure metric exhibited a significantly higher power across all values of α for $V = 1$, at the 0.01 level of significance for $\alpha = 0.03$ and $\alpha = 0.05$ and at the 0.05 level of significance at $\alpha = 0.01$. At $V = 0.3$, the SHEDS-PM exposure metric exhibited a significantly higher power at the 0.01 level of significance for $\alpha = 0.03$ and $\alpha = 0.05$. It can be seen in Figure 8 that with the exception of the case $V = 1.0$ and $\alpha = 0.3$, for all values of V and α considered, the SHEDS-PM personal exposure metric exhibits higher power than the AQS exposure metric. Other possible values for E_{sd} were considered as well. $E_{sd} = 0.2$ (not shown) showed no significant difference in the power of detecting a non-zero effect of $PM_{2.5}$ between the AQS and SHEDS-PM exposure metrics. As seen in Table 1, $E_{sd} = 0.4$ showed a significant difference at the 0.05 level in the power of detecting a non-zero effect of $PM_{2.5}$ between the AQS and SHEDS-PM exposure metrics. The AQS metric exhibited a significantly higher power at $\alpha = 0.01$, there was no difference at $\alpha = 0.03$, and SHEDS-PM exhibited a significantly higher power at $\alpha = 0.05$ for $E_{sd} = 0.4$. It is important to point out that in most cases the powers are relatively similar, and that in reality, SHEDS-PM will not summarize the population exposure distribution as it does for our synthetic data. This simulation simply provides an illustration of the statistical properties of the SHEDS-PM output.

Incorporating personal exposure increases our power in detecting risk for an increase in expected number of hospital admissions due to fine particulate matter. It has increased power across increasing magnitudes of relative risk.

5 Application Study

We consider an analysis of $PM_{2.5}$ metrics - monitoring data from AQS, modeled CMAQ output from the EPA, and personal exposure with SHEDS-PM - with health data from NMMAPS. The response considered in this application is respiratory disease emergency hospital admissions for Medicare patients. First we consider the ratio of individual exposure to the input ambient concentration, E_a/C, as an exposure metric. E_a/C is often used to study the output of exposure simulators (Ozkaynak et al [38]).

5.1 Ea/C Analysis

We consider E_a/C over time, to investigate its temporal properties. Figure 3 shows that E_a/C is relatively stable over time, with some seasonal fluctuations. This is one of the reasons that we use E_a as the predictor for the analysis in this case, as individual variation might be more informative than the concentration ratios.

E_a/C is important to consider because it portrays information about individual sources of variability such as housing type and activity patterns. Mean E_a/C is hypothesized to vary by season, as depicted in Figure 3, as well as by geographic location (Cao and Frey [29], [39]). Though in this case there is little differentiation in the quantiles for E_a/C across the three counties, as seen in Table 2. Figure 3 shows E_a/C ratios over time for 2003. There is no evidence of a significant linear trend though there is a clear seasonal pattern.

Table 2 shows E_a/C ratios for 2002-2006 overall and by county. January 2003 and July 2003 are also shown to represent both a "cold" and "warm" month for comparison. Figure 4 shows E_a/C ratios for a representative day for the "cold" and "warm" seasons: January 15, 2003 (a) and July 14, 2003 (b) respectively. There is evidence of a county effect for the E_a/C ratio, with New York County showing a more skewed distribution with significant positive mass in the upper tail. Figure 4 shows the relative variation across the population during a single day of exposure, versus the single exposure value for the AQS metric represented by the vertical line. The goal of SHEDS-PM is to model the variation in the distribution of possible exposure values across different members of the population. Figure 4 shows the amount of information contained in SHEDS-PM exposure metric relative to the static AQS metric.

Since the E_a/C are not exhibiting a linear temporal trend, E_a - i.e. ambient concentration - is used as the personal exposure metric. Seasonality is captured in the model via the linear and quadratic terms in time, and the spline fit for temperature.

6 Results

Results for the effect of fine particulate matter exposure on emergency hospital admissions for respiratory cases showed a positive association between increased exposure and number of admissions for all metrics. We also considered an analysis with cardiovascular emergency hospital admissions as the response. This study did not show a consistent effect of fine particulate matter on cardiovascular admissions. The sign for the estimated effect of $PM_{2.5}$ on cardiovascular admissions was negative but not significant for AQS. In this study we focus on the details of the respiratory outcomes, in order to compare the effectiveness of SHEDS-PM as a metric in contrast with the more widely used and studied AQS.

6.1 Non-Individual Exposure Models: AQS and CMAQ

This section details the results for the AQS and CMAQ pollutant exposure surrogates for the non-individual exposure models. Table (3) shows the posterior coefficient estimates for both ozone and $PM_{2.5}$ and their corresponding 95% credible intervals for the AQS and CMAQ exposure models.

We used normal priors with mean 0 and variance 100 as uninformative priors for the $PM_{2.5}$ exposure metric coefficient parameter to allow the data to inform the posterior. Both the AQS and CMAQ exposure metrics exhibit a positive coefficient for $PM_{2.5}$, indicating that the relative risk for emergency hospital admissions for respiratory disease increases with increased fine particulate matter exposure. For AQS, the posterior mean of β_{PM} is 0.0179 with a 95% posterior credible interval of $(0.0008, 0.0350)$, which corresponds to an increased relative risk of approximately 1.8% ($e^{0.0179} = 1.018$) for emergency respiratory hospital admissions. This corresponds to an approximate increase of 1.8 admissions per 100, with a posterior 95% credible interval 0.08 to 3.3, for each one standard deviation increase fine particulate matter ($PM_{2.5}$) on a given day. For CMAQ, the posterior mean of β_{PM} is 0.0224 with a 95% posterior credible interval of $(0.0128, 0.0321)$, which corresponds to an increased relative risk of approximately 2.2%, i.e. an approximate increase of 2.2 admissions per 100 for each one standard deviation increase in fine particulate matter ($PM_{2.5}$) on a given day, with a posterior 95% credible interval (1.3, 3.3). It is important to point out that CMAQ results in more precise estimates than AQS, as evidenced by the smaller credible intervals and posterior standard deviation.

6.2 Individual Exposure Models: AQS and CMAQ

Table (4) shows the posterior coefficient estimates for both ozone and $PM_{2.5}$ and their corresponding 95% credible intervals for the SHEDS-PM individual exposure model. Table (4) also showcases the posterior coefficient estimates and corresponding credible intervals for the full model considered, including the linear, quadratic, and spline terms for the additional confounders. SHEDS-PM is exhibiting a positive coefficient for $PM_{2.5}$, indicating that the relative risk for emergency hospital admissions for respiratory disease increases with increased levels of individual exposure to fine particulate matter. For SHEDS-PM, the posterior mean of α_{PM} is 0.0233 with a 95% posterior credible interval of $(0.0135, 0.0332)$, which corresponds to an increased relative risk of approximately 2.4% for emergency respiratory hospital admissions. This corresponds to an approximate increase of 2.4 admissions per 100, with a 95% posterior credible interval of (1.4, 3.4) for each one standard deviation increase in fine particulate matter ($PM_{2.5}$) on a given day.

SHEDS-PM results in more precise estimates than AQS, as shown by the smaller credible intervals, and is comparable to CMAQ in this regard. The uncertainty associated with the SHEDS-PM coefficient is less than that of AQS, showing a 43% reduction in uncertainty estimates. The uncertainty associated with SHEDS-PM is comparable to that of CMAQ.

Figure (5) shows the posterior distribution of $PM_{2.5}$ coefficient estimates for the AQS, CMAQ, and SHEDS-PM metrics. The uncertainty associated with the AQS coefficient estimates is higher than that of CMAQ and SHEDS-PM. In addition, the $PM_{2.5}$ coefficient posterior estimates for CMAQ and SHEDS-PM are consistent with each other with regards to the posterior mean (0.0224 for CMAQ and 0.0233 for SHEDS-PM respectively). This indicates that the additional information contained in the individual

exposure metric of SHEDS-PM may provide more precise estimates of the effect of $PM_{2.5}$.

Sensitivity analysis detailed in Section 6.3 indicates that confounding factors such as temperature and time were satisfactorily addressed. The simulation study shows that SHEDS-PM exhibits a higher power for detecting an increase in relative risk than AQS and CMAQ, with power increasing as a function of the true magnitude of the relative risk coefficient. Several reasonable values for the prior variance were considered to test prior robustness with similar results.

Results for cardiovascular admissions, while not detailed here, were similar in terms of the comparison between metrics. The estimated effects of $PM_{2.5}$ on cardiovascular emergency admissions were similiar for CMAQ and SHEDS-PM, and the corresponding uncertainty estimates were more precise for SHEDS-PM compared to AQS.

6.3 Sensitivity Analysis

A sensitivity analysis was run to determine the appropriate degrees of freedom for the spline fits in the Poisson model. Splines were fit for the following confounders using the function *ns* in the R-package *gam* [40]. Degrees of freedom for the spline fits were selected using a sensitivity analysis on the coefficient of the covariate of interest, $PM_{2.5}$, in the model. Figure (6) shows the sensitivity the coefficient of $PM_{2.5}$ to the spline fit for the possible values 1-10 for degrees of freedom for AQS and CMAQ for respiratory admissions. The degrees of freedom selected, according to the stabilization of the coefficient estimate, were $d_1 = 3$ for temperature and $d_2 = 3$ for relative humidity. This is in relative agreement with commonly used literature values [34]. Though Peng et al (2006) [34] utilized higher degrees of freedom for the spline fits, their analysis concerned PM_{10} over a 13 year period, where here $PM_{2.5}$ is analyzed over a 5 year period.

Preliminary analysis showed significance for the linear and quadratic terms in time for respiratory disease response. Several models were considered including spline fits in time, linear and quadratic terms in temperature as well as relative humidity, and various lag values for temperature, ozone, and $PM_{2.5}$. The spline fits in time were not significant, possibly due to the splines in the other covariates capturing of portion of the temporal trend, including the spline and linear fit in termperature. The quadratic time fit was significant to capture the temporal trend given information in the other spline terms. The base model was selected using the significance of the terms in the model as well as overall model AIC values. Exploratory data analysis showed very mild overdispersion, with values of the dispersion parameter estimated between 1.05 and 1.52, thus the standard Poisson model is appropriate. Figure (7) shows the amount of trend captured in the model for emergency respiratory admissions. The blue shows the effect of the confounders on explaining emergency hospital respiratory admissions and the red indicates the added effect of $PM_{2.5}$, utilizing a generalized linear model fit for exploratory data analysis.

7 Discussion

In this work we study the impact of using a population exposure model, SHEDS-PM, as a metric to characterize particulate matter in studying the relative risk for emergency hospital admissions. SHEDS-PM uses information about demographics and activity patterns of the population of interests as well as a characterization of the potential indoor exposure resulting in a more complete description of the population exposure.

Our results indicate that SHEDS-PM provides approximately the same increase in relative risk associated with emergency respiratory admissions as using a chemistry model, CMAQ, or monitoring data, AQS, as exposure metrics. However, SHEDS-PM and CMAQ both bring additional information which helps to reduce the uncertainly in our estimated risk by approximately half. The exposure models SHEDS-PM and CMAQ have errors and sources of uncertainty, and further evaluation of these models is recommended, since this exposure model error could result in a bias in the estimated risk. SHEDS-PM provides additional power over AQS in detecting a positive effect on relative risk for emergency hospital admissions associated with $PM_{2.5}$ exposure.

SHEDS-PM is a very useful model for characterizing population exposure to $PM_{2.5}$. In comparison to CMAQ, SHEDS-PM does not provide additional information for the characterization of relative risk with regards to exposure. However, while CMAQ can provide output at a very high resolution, it is specific to the CMAQ grid cell location, and does not account for population variability introduced by possible movement across grid cells. SHEDS-PM provides a metric capable of capturing this variability, as it is based on human demographics and activity patterns and time spent in various microenvironments. The additional information available in the personal exposure metric provides a more complete description of population exposure at the county level, as in this study, as individuals are not static within one grid cell. There could possibly be an additive exposure effect that could be represented by this variation in activity patterns that is possibly being absorbed into the information provided by the other model covariates. In addition, if health data was available resolved at a finer geographical scale, SHEDS-PM could provide more realistic spatial variation in daily exposures for the estimation of health effects at the census tract resolution.

In order to make a direct comparison to the often used exposure surrogates AQS and CMAQ, only the ambient individual exposure through SHEDS-PM was considered. SHEDS-PM also provides information about non-ambient individual exposure, such as exposure through smoking or cooking. There was evidence of a county effect in the personal exposure distribution. As the focus of this study was to make an initial comparison of exposure metrics and modeling a county effect would present an interesting challenge in and of itself, capturing a county effect is left for future work in order to keep focus on the exposure metrics under consideration. In addition, AQS and CMAQ contain information about speciated particulate matter, including nitrate, sulfate, elemental carbon, and organic carbon. Adding these additional covariates into a modeling scheme for adverse human health effects greatly increases the complexity. An area of current study is the

consideration of model selection techniques and controlling for multi-collinearity in the presence of these additional covariates.

8 Acknowledgements

The authors would like to acknowledge Howard Chang at Emory University and Laura Boehm at North Carolina State University for their contributions regarding modeling methodology and data management. We also acknowledge Drs. Valeria Garcia and Halûk Özkaynak with the U.S. Environmental Protection Agency's National Exposure Research Laboratory, for providing the CMAQ output data and for guidance with the application of the SHEDS-PM model for New York City, respectively. Elizabeth Mannshardt was supported as a Post Doctoral Research Scholar through the National Science Foundation's Collaborative Research: RNMS Statistical Methods for Atmospheric and Oceanic Sciences under Grant No. DMS-1107046. Katarina Sucic was supported in part through Havard University's Statistical Methods for Population Health Research on Chemical Mixtures, Grant No. 114346-5053742. Montserrat Fuentes was sponsored in part by the National Institutes of Health under Grant No. 2R01ES014843-04A1. Francesca Dominici was supported in part by the National Institutes for Health, Grant No. R01ES019560 and the Environmental Protection Agency, Grant No. RD83479801. H. Christopher Frey and Wan Jiao were sponsored by the National Institutes of Health under Grant No. 1R01ES014843-01A2 and U.S. EPA STAR Grant RD 83386301. This paper has not been subject to review by the NSF, NIH or EPA, and the authors are solely responsible for its content.

References

[1] Dominici F, Peng RD, and Bell ML. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*, pages 1127–1134, 2006.

[2] Pope CA III and Dockery DW. Health effects of fine particulate air pollution: lines that connect. *J Air Waste Manag Assoc.*, pages 709–742, 2006.

[3] Bell M, Ebisu K, and an Dominici F. Peng R. Adverse health effects of particulate air pollution: modification by air conditioning. *Epidemiology*, pages 682–686, 2009.

[4] Ostro B, Roth L, Malig B, and M. Marty. The effects of fine particle components on respiratory hospital admissions in children. *Environ Health Perspect*, pages 475–480, 2009.

[5] Ostro BD, Broadwin R, and Lipsett MJ. Coarse and fine particles and daily mortality in the Coachella Valley: A follow-up study. *Journal of Exposure Analysis and Environmental Epidemiology*, pages 412–419, 2000.

[6] Braga ALF, Zanobetti A, and Schwartz J. The lag structure between particulate air pollution and respiratory and cardiovascular deaths in ten US cities. *Journal of Occupational and Environmental Medicine*, pages 927–933, 2001.

[7] Holloman CH, Bortnick SM, Morara M, Strauss WJ, and Calder CA. A Bayesian hierarchical approach for relating $PM_{2.5}$ exposure to cardiovascular mortality in North Carolina. *Environmental Health Perspectives*, pages 1282–1288, 2004.

[8] Hoek G, Brunekreef B, Fisher P, and Wijnen JV. The association between air pollution and heart failure, arrythmia, embolism, thromobosis, and other cardiovascular causes of death in a time series study. *Epidemiology*, pages 355–357, 2001.

[9] Pope CA, Dockery DW, and Schwartz J. Review of epidemiological evidence of health-effects of particulate air-pollution. *Inhalation Toxicology*, pages 1–18, 1995.

[10] Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris Jr BG, and Speize FE. An association between air pollution and mortality in six US cities. *New England Journal of Medicine*, pages 1753–1759, 1993.

[11] Dockery D and Pope CA. Acute respiratory effects of particulate air pollution. *Annu Rev Public Health*, pages 107–132, 1994.

[12] Fuentes M, Song H-R, Ghosh SK, Holland DM, and Davis JM. Spatial association between speciated fine particles and mortality. *Biometrics*, pages 855–863, 2006.

[13] `http://www.epa.gov/pm`.

[14] Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, Dockery D, Schwartz J, and Al. Zanobetti. *The National Morbidity, Mortality, and Air Pollution Study (HEI project no. 96-7): morbidity and mortality from air pollution in the United States*. Cambridge, MA: Health Effects Institute, 2000.

[15] Byun DJ and Schere KL. Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (cmaq) modeling system. *Appl Mech Rev.*, pages 51–77, 2006.

[16] Calder CA, Holloman CH, Bortnick SM, Strauss WJ, and Morara M. Relating ambient particulate matter concentration levels to mortality using an exposure simulator. *Journal of the American Statistical Association*, pages 137–148, 2008.

[17] Berrocal VJ, Gelfand AE, Holland DM, Burke J, and Miranda ML. On the use of a $PM_{2.5}$ exposure simulator to explain birthweight. *Environmetrics*, pages 553–571.

[18] Blangiardo M, Hansell A, and Richardson S. A bayesian model of time activity data to investigate health effect of air pollution in time series studies. *Atmospheric Environment*, pages 379–386, 2011.

[19] Reich BJ, Fuentes M, and Burke J. Analysis of the effects of ultrafine particulate matter while accounting for human exposure. *Environmetrics*, pages 131–146, 2009.

[20] Peng RD, Chang HH, Bell ML, McDermott A, Seger SL, Samet JM, and Dominici F. Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among medicare patients. *American Medical Association*, pages 2172–2179, 2008.

[21] Binkowski FS and Roselle SJ. Models-3 community multiscale air quality (CMAQ) model aerosol component, 1. Model description. *J. Geophys. Res.*, 2003.

[22] Fuentes M and Raftery A. Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models. *Biometrics*, pages 36–45, 2005.

[23] `http://www.epa.gov/AMD/CMAQ/cmaq_model.html`.

[24] Chang HH, Reich BJ, and ML Miranda. Time-to-event analysis of fine particle air pollution and preterm birth: results from North Carolina, 2001-2005. *American Journal of Epidemiology*, pages 91–98, 2012.

[25] Bravo MA, Fuentes M, Zhang Y, Burr MJ, and Bell ML. Comparison of exposure estimation methods for air pollutants: Ambient monitoring data and regional air quality simulation. *Environmental Research*, pages 1–10, 2012.

[26] Burke JM and Vedamtham R. Stochastic human exposure and dose simulation for particulate matter (SHED-PM) version 3.5 user guide. *US Environmental Protection Agency*, 2009.

[27] `http://www.epa.gov/chadnet1/`.

[28] Kinney PL, Chllrud SN, Sax S, Ross JM, , Macintosh D, Myatt TA, and Spengler JD. Toxic exposure assessment: A columbia-harvard (TEACH) study (the New York City Report). *NUATRC Research Report Number 3*, 2005.

[29] Cao Y and Frey HC. Geographic differences in inter-individual variability of human exposure to fine particulate matter. *Atmospheric Environment*, pages 5684–5691, 2011.

[30] Jones R, zkaynak H, Nayak S, Garcia V, Hwang S-A, and Linn S. Associations between summertime ambient pollutants and respiratory morbidity in NYC: comparison of ambient concentrations versus predicted exposures. *Journal of Exposure Science and Environmental Epidemiology*, 2012: To be submitted.

[31] Cao Y and Frey HC. Assessment of interindividual and geographic variability in human exposure to fine particulate matter in environmental tobacco smoke. *Risk Analysis*, 31(4):578–591, 2010.

[32] Dominici F, Zeger SL, and Samet JM. A measurement error model for time-series studies of air pollution and mortality. *Biostatistics*, pages 157–175, 2000.

[33] Dominici F. Invited commentary: air pollution and health- what can we learn from a hierarchical approach? *Am J Epidemiol.*, pages 11–15, 2002.

[34] Peng RD and Louis TA. Dominici F. Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society Series A*, pages 179–203, 2006.

[35] Schwartz J. Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Can. J. Statist*, pages 471–488, 1994.

[36] Choi J, Fuentes M, and Reich BJ. Spatial-temporal association between fine particulate matter and daily mortality. *Computational Statistics and Data Analysis*, pages 2989–3000, 2009.

[37] Berger JO. *Statistical Decision Theory and Bayesian Analysis*. Springer–Verlag, New York, 1985.

[38] Ozkaynak H, Frey HC, Burke J, and RW Pinder. Analysis of coupled model uncertainties in source to dose modeling of human exposures to ambient air pollution: a PM2.5 case-study. *Atmospheric Environment*, pages 1641–1649, March 2009.

[39] Sarnat R, Wilson W, Strand M, Brook J, Wyzga R, and Lumley T.

[40] `http://cran.r-project.org/web/packages/gam/gam.pdf`.

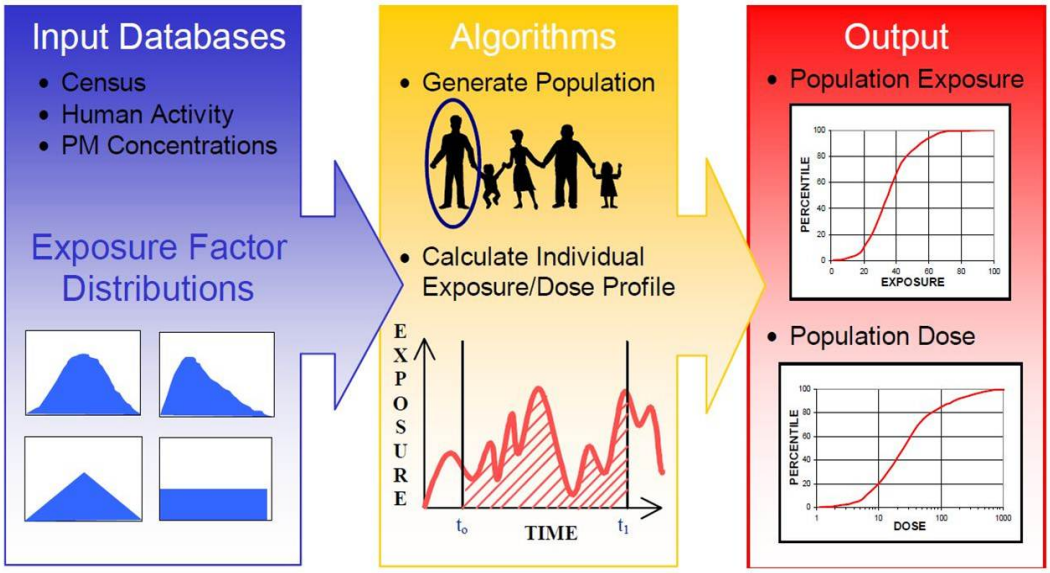


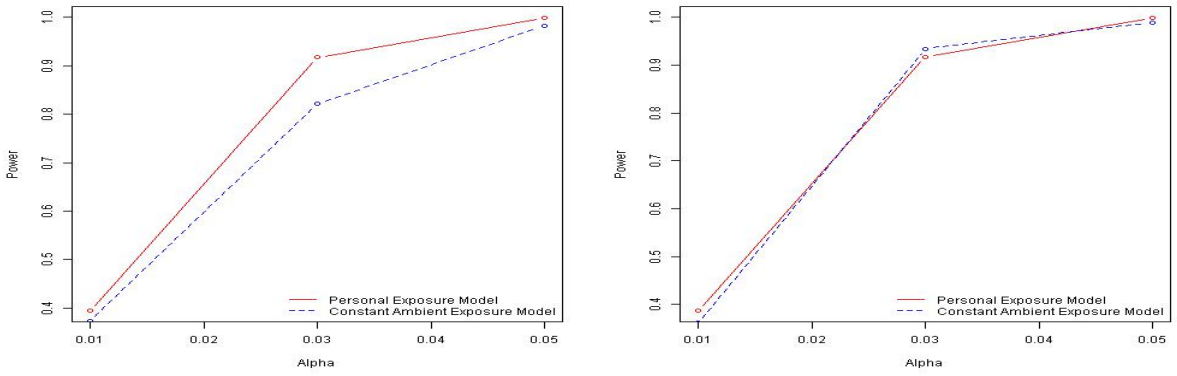
Figure 1: Input and output for SHEDS-PM Model [26]. Used with permission.

E_{sd}	V	α=0.01		α=0.03		α=0.05	
		PE	Amb	PE	Amb	PE	Amb
0.4	0.3	0.270	0.293*	0.649	0.633	0.929*	0.915
		(0.011)	(0.011)	(0.012)	(0.012)	(0.006)	(0.007)
	1	0.265	0.294*	0.665	0.648	0.929*	0.915
		(0.011)	(0.011)	(0.012)	(0.012)	(0.006)	(0.007)
0.88	0.3	0.395	0.373	**0.917**	0.821	**0.998**	0.982
		(0.012)	(0.012)	(0.007)	(0.010)	(0.001)	(0.003)
	1	0.387*	0.362	**0.934**	0.827	**1.000**	0.989
		(0.012)	(0.012)	(0.006)	(0.009)	(0.000)	(0.003)

Table 1: Table of Power for estimated personal exposure versus AQS using $N_1 = 1623$ where PE represents the inclusion of personal exposure versus overall ambient concentration, Amb. Standard errors are in parenthesis, * indicates significance at the 0.05 level, and **bold** indicates significance at the 0.01 level for testing that the power of personal exposure metric is equal to the power of the ambient exposure metric.

	Overall			Jan 2003			Jul 2003		
Percentiles	25^{th}	50^{th}	95^{th}	25^{th}	50^{th}	95^{th}	25^{th}	50^{th}	95^{th}
All	0.559	0.656	0.865	0.535	0.629	0.837	0.620	0.704	0.882
Bronx	0.558	0.656	0.870	0.536	0.630	0.843	0.619	0.704	0.885
NY	0.558	0.655	0.859	0.534	0.629	0.834	0.620	0.704	0.879
Queens	0.560	0.657	0.866	0.535	0.629	0.837	0.620	0.705	0.883

Table 2: E_a/C ratios for 2002-2006



(a) (b)

Figure 2: Power across $\alpha = 0.01, 0.03, 0.05$. V fixed at 0.3 (a) and 1.0 (b), with E_{sd} at 0.88. The red solid line represents the personal expsosure metric and the blue dashed line represents the AQS exposure metric.

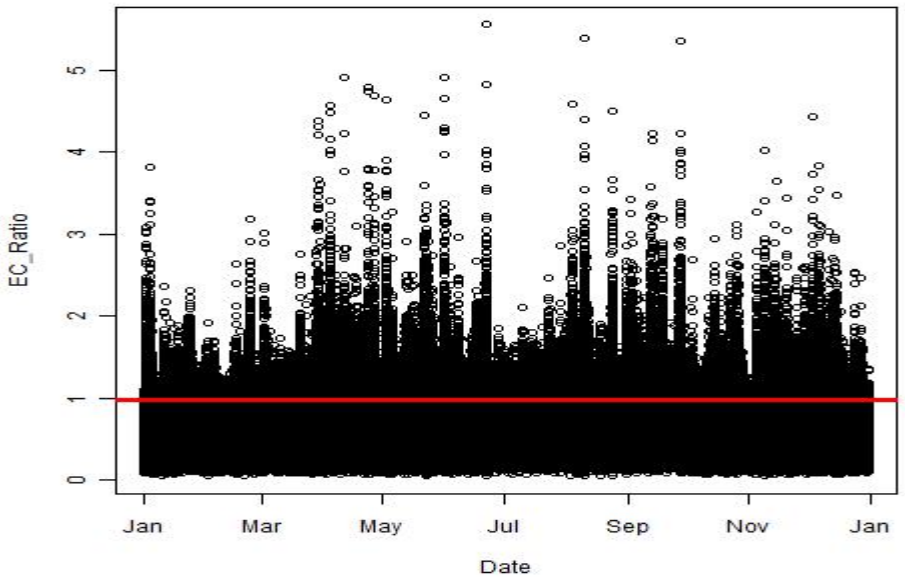
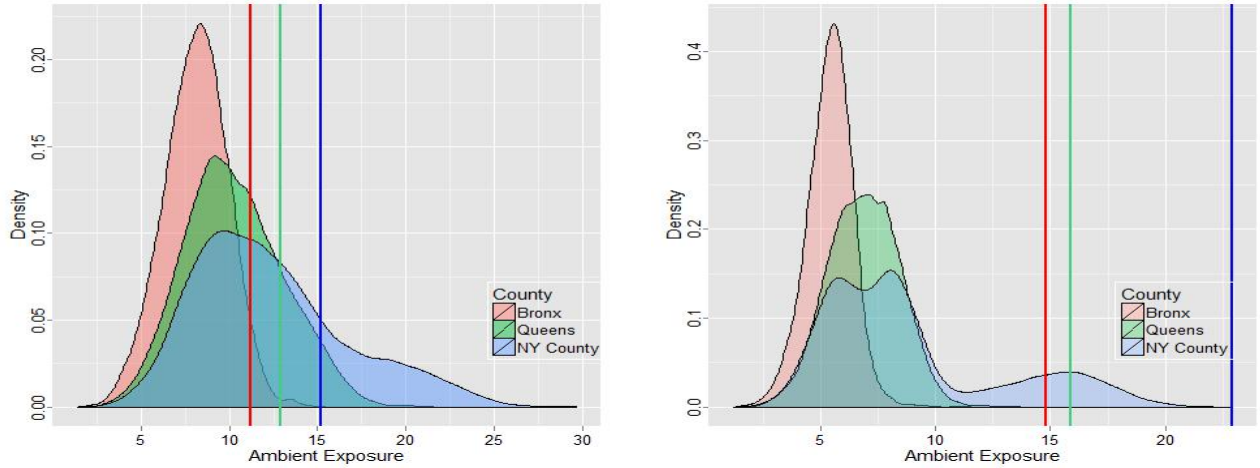


Figure 3: E_a/C ratio distribution across all counties for 2003. Red line at 0.993 indicates the 99^{th} percentile.



(a) (b)

Figure 4: SHEDS-PM distribution for Jan 15 2003 (a) and Jul 14, 2003 (b) across the three counties. Bronx county is in red, Queens in green, and New York County in blue. Vertical lines represent AQS input concentration values.

Covariate	Posterior Mean	Posterior Std Dev	2.5^{th} percentile	97.5^{th} percentile
AQS	0.0179	0.0088	0.0008	0.0350
CMAQ	0.0224	0.0049	0.0128	0.0321

Table 3: AQS and CMAQ posterior distribution of the effect of ambient $PM_{2.5}$ on emergency respiratory admissions

Covariate	Posterior Mean	Standard deviation	2.5^{th} percentile	97.5^{th} percentile
Intercept	-0.3177	0.0609	-0.4493	-0.2097
t^2	0.0001	0.0000	0.0001	0.0002
t^2	-0.0000	0.0000	-0.0000	-0.0000
temp	-0.0074	0.0025	-0.0118	-0.0025
temp.Sp1	0.0282	0.1241	-0.2096	0.2582
temp.Sp2	1.0405	0.2614	0.5759	1.5431
temp.Sp3	0.3104	0.1641	0.0025	0.6077
hum.Sp1	0.0331	0.0218	-0.0091	0.0760
hum.Sp2	0.0133	0.0753	-0.1432	0.1643
hum.Sp3	0.0176	0.0247	-0.0309	0.0663
Sunday	-0.0460	0.0169	-0.0791	-0.0128
Monday	0.1959	0.0162	0.1646	0.2276
Tuesday	0.1151	0.0166	0.0829	0.1479
Wednesday	0.1193	0.0162	0.0883	0.1511
Thursday	0.0874	0.0163	0.0557	0.1202
Friday	0.1383	0.0163	0.1065	0.1705
Lag1Ozone	-0.0418	0.0063	-0.0539	-0.0294
Lag1$PM_{2.5}$	0.0233	0.0050	0.0135	0.0332

Table 4: SHEDS posterior distribution for the effect of ambient $PM_{2.5}$ and confounding covariates on emergency respiratory admissions

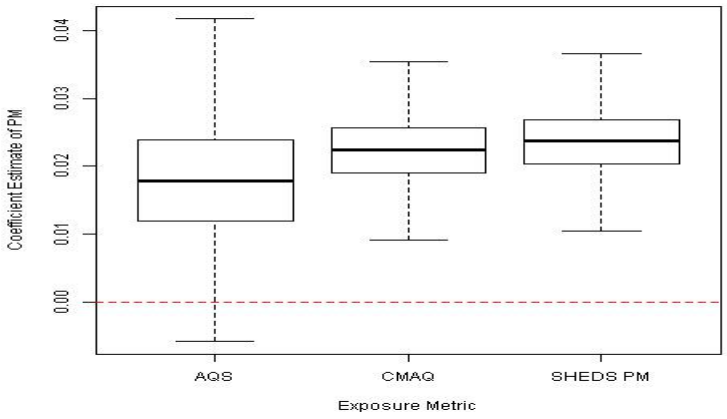


Figure 5: Posterior distribution of $PM_{2.5}$ coefficients estimates for emergency respiratory admissions

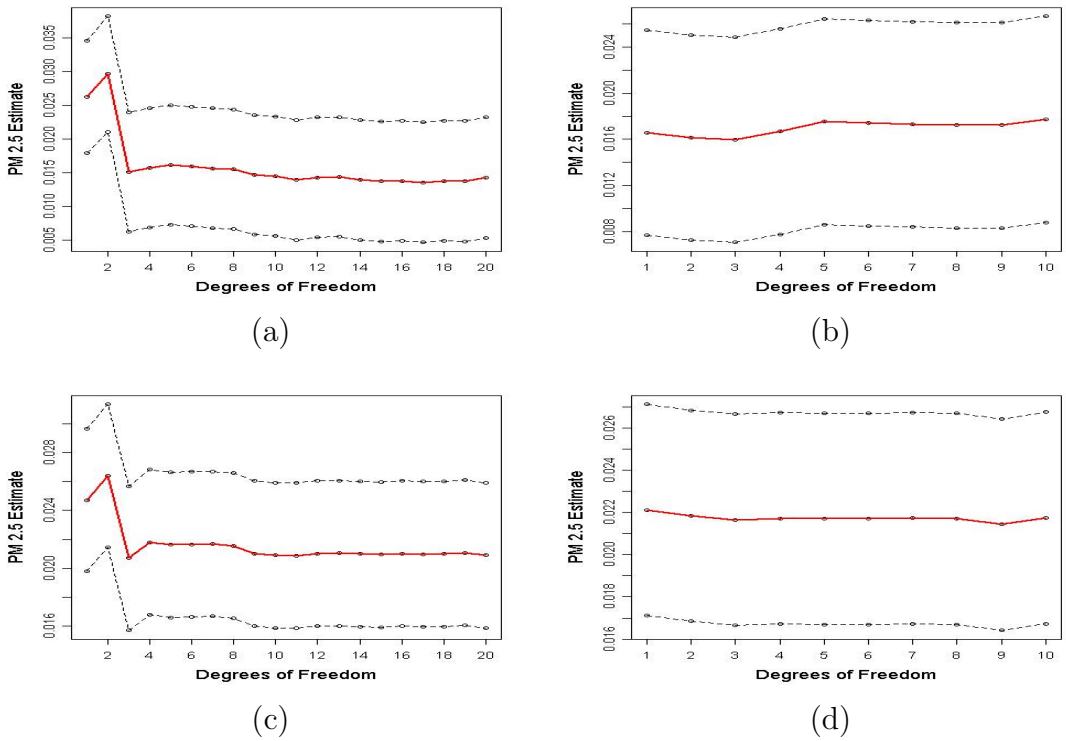
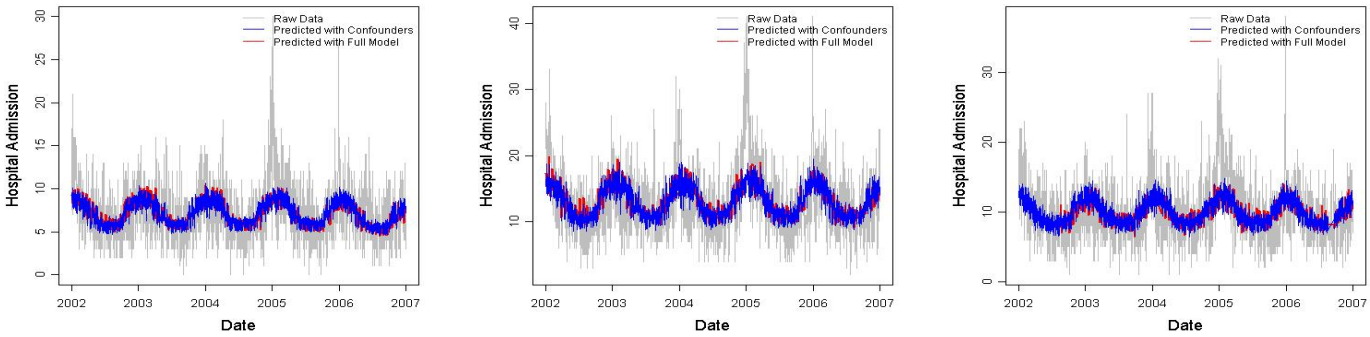


Figure 6: Spline sensitivity analysis over all counties for temperature (a) and relative humidity (b) on PM coefficient for AQS; temperature (c) and relative humidity (d) on PM coefficient for CMAQ.



(a) (b) (c)

Figure 7: Base Model AQS Fits: (a) Bronx, (b) Queens, & (c) New York County. Utilizing a generalized linear model fit, the blue lines show the effect of the confounders on emergency respiratory admissions, and the red indicates the added effect of $PM_{2.5}$.