Effect of Outdoor Airborne Particulate Matter on Daily Death Counts

Patricia Styer,[1] Nancy McMillan,[1] Feng Gao,[2] Jerry Davis,[1,3] and Jerome Sacks[1]

[1]National Institute of Statistical Sciences, Research Triangle Park, NC 27709-4162 USA; [2]Battelle Pacific Northwest Laboratories, Richland, WA 99352 USA; [3]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695-8208 USA

To investigate the possible relationship between airborne particulate matter and mortality, we developed regression models of daily mortality counts using meteorological covariates and measures of outdoor PM_{10}. Our analyses included data from Cook County, Illinois, and Salt Lake County, Utah. We found no evidence that particulate matter ≤10 µm (PM_{10}) contributes to excess mortality in Salt Lake County, Utah. In Cook County, Illinois, we found evidence of a positive PM_{10} effect in spring and autumn, but not in winter and summer. We conclude that the reported effects of particulates on mortality are unconfirmed. *Key words*: causal inference, model selection, observational data, PM_{10}, Poisson regression, semi-parametric modeling. *Environ Health Perspect* 103:490–497(1995)

To determine if airborne particulates contribute to excess mortality, researchers have adopted multiple regression techniques to measure the effects of particulates on daily death counts (*1,2*). Other factors, such as extreme temperatures, can affect mortality, and regression techniques are used to adjust for these other known influences. Though many factors could be involved, research has generally limited attention to meteorological sources such as temperature and humidity. In some cases, other air pollution measures such as sulfur dioxide and ozone are included. The regression coefficient corresponding to a measure of particulate level is then interpreted as the effect of particulate pollution on mortality, accounting for stress from the other influences. If this coefficient is a statistically significant positive number, the conclusion is that mortality increases with increasing levels of particulates. This association is then elevated to a causal interpretation: particulates cause death, and researchers estimate that soot at levels well below the maximum set by federal law "kills up to 60,000 in U.S. each year" (*3,4*), and similar calculations "put the annual toll in England and Wales at 10,000" (*5*).

Studies vary as to the particulate measures used and the locations analyzed. In the analyses presented here, we used PM_{10}, which specifies particulate matter with an aerodynamic diameter ≤10 µm (*6*). The current U.S. EPA standard is based on this measure. The locations we analyzed, Cook County, Illinois, and Salt Lake County, Utah, both have relatively long records of PM_{10} monitoring. The monitoring data are discussed in more detail in Methods.

The data used in the analyses (meteorological conditions, particulate levels, death counts) are observational; that is, data that are measured and recorded without control or intervention by researchers. Deducing causal relationships from observational data is perilous. A practical approach described by Mosteller and Tukey (*7*) involves considerations beyond regression analysis. In particular, consideration should be given to whether the association between particulate levels and mortality is consistent across "settings," whether there are plausible common causes for elevated particulate levels and mortality, and whether the derived models reflect reasonable physical relationships.

There is a high degree of association of PM_{10} with meteorology, and a high degree of association of mortality with weather. For example, in the summer in Cook County the correlation coefficient between the daily average of PM_{10} and daily mean temperature is 0.52 and the correlation between daily elderly (age 65 or older) mortality and mean temperature is 0.25. The confounding effects of weather as a partial cause of both particulate levels and mortality may not be controllable by standard regression methods; the appearance of an effect for particulates, i.e., a positive coefficient for the PM_{10} term, may, as a result, be spurious (see Appendix B). We have not addressed the issue of errors in variables, which can also be a cause for spurious relationships. The concern about errors in variables arises from the differences between measured PM_{10} and the actual PM_{10} exposure experienced by the population. PM_{10} measurements are taken outdoors, but people tend to spend most of their time indoors, especially the sick and elderly who are believed to be the most vulnerable. Similarly, the meteorological covariates we include represent outdoor conditions. And again, when explanatory variables are measured with error, the result is not necessarily attenuation of the regression surface. In multiple regression, the result can be an artificial increase in the magnitude of the estimated coefficients.

The results for Cook County and Salt Lake County show that the appearance and size of a PM_{10} effect is quite sensitive to model specification. In particular, the treatment of season affects the estimates of the PM_{10} effect. In Cook County, we found a significant interaction between the time of year and PM_{10}. Using a standard Poisson regression model, we found that PM_{10} appears to be significantly associated with mortality in the spring and fall, but not in the winter and summer. Using a semi-parametric model (Appendix A), we found that only the months of May and September exhibit a particulate effect. In Salt Lake County, the semi-parametric model suggests a similarly isolated PM_{10} effect limited to the month of June, but we found no evidence of a PM_{10} effect in any model using Poisson regression. Hence, we conclude there is no evidence of a consistent association between particulates and mortality.

Several studies carried on at various locations in the United States have reported small yearly increases in mortality resulting from increases in particulates. In our Cook County analyses, the effect of PM_{10} in the spring and fall induces a similar positive yearly increase in mortality from increases in particulates, but the increase is from one-half to one-third the size usually reported in other studies depending on the analyses performed. In Salt Lake County, the size of the yearly effect is far smaller and statistically insignificant. What remains unexplained is why, in Cook County, effects should appear in the spring but not in the summer, and in the fall but not in the winter. Neither is it clear why the effect of particulates on mortality should not appear in any season in Salt Lake County.

The appearance of a PM_{10} effect in the spring and fall in Cook County led to the speculation that pollen may be implicated, but no such evidence was found using pollen data monitored in the city of Chicago, the major population component of Cook County. Other analyses carried out for the fall season in Cook County on different subgroups of the population produced no definitive differences among subgroups.

The inconsistency of the regression analyses, the unresolved status of plausible common causes of particulate levels and mortality, the confounding effects of weather, and the unavailability of plausible biophysical mechanisms to explain the

Address correpondence to J. Sacks, National Institute of Statistical Sciences, PO Box 14162, Research Triangle Park, NC 27709-4162 USA.
This research was supported in part by the U.S. Environmental Protection Agency under cooperative agreement CR819638-01-0 and by a National Science Foundation Grant (DMS-9208758).
Received 27 October 1994; accepted 15 February 1995.

empirical analyses prevent us from concluding that there is an effect between "today's" mortality and "yesterday's" particulates. The question appears to be unresolved.

Methods

Data

The data used for the statistical studies have three main components: mortality counts, particulate levels, and meteorology. The sources of the data are described in this section along with some summary statistics.

Mortality data. Daily death counts for the period 1985 through 1990 came from death certificate records for Cook and Salt Lake County residents, collected by the National Center for Health Statistics, and made available to us by John Creason, EPA. Although mortality data are available for longer periods, PM_{10} data are unavailable before 1985. Each death record contains a cause of death code and some basic demographic information. In compiling daily death counts, we excluded all deaths from accidental causes, as well as deaths of county residents occurring in other locations. We refer to the remaining number of deaths as total deaths. The main analyses were performed with total deaths among the population aged 65 or older (elderly deaths). We carried out additional analyses for total deaths, unrestricted by age, for deaths classified by specific causes, and for selected population subgroups such as elderly blacks and elderly males. We classified the disease-specific causes of death by the International Classification of Diseases (ICD) codes that appear on the mortality records. We adopted the classification scheme detailed in Fairley (*8*), extracting cancer deaths (ICD categories 140–209), circulatory deaths (ICD categories 390–459), and respiratory deaths (ICD categories 11, 35, 472–519,710.0, 710.2,710.4).

In Cook County, there was an average of 117 nonaccidental deaths per day for all ages. Among residents aged 65 and over, there was an average of 83 deaths per day. Death counts vary by time of year, with higher numbers in winter and fewer deaths in summer. In Salt Lake County, there was an average of 9 nonaccidental deaths for all ages and 7 nonaccidental deaths for residents 65 and over. As in Cook County, there are slightly more deaths in the winter. Table 1 displays some summary statistics for both Cook County and Salt Lake County mortality.

Particulate data. In current monitoring efforts, particulates are measured throughout the United States. There are both 24-hr and annual ambient air quality standards for particulate matter (*6*). In the first case, the standard is attained when the expected number of days per calendar year with a 24-hr average concentration above 150 $\mu g/m^3$ is equal to or less than one. In the second case, the standard is attained when the expected annual arithmetic mean concentration is less than or equal to 50 $\mu g/m^3$. To comply with these standards, it is sufficient to collect samples from each monitoring site only once every 6 days, though there are a few locations with monitors that operate on a daily basis. For Cook County, the particulate data come from a network of PM_{10} monitors reported in the EPA Aerometric Information Retrieval System (AIRS) for the period 1985 through 1990. During this time, there were 20 separate monitors in operation, though several monitors were operated for only a brief period of time. The Cook County network includes one daily station where PM_{10} samples are collected on a daily basis. The remaining stations collected samples every sixth day. The daily station observations are frequently missing, with 69% of the values recorded once the monitoring station began operation in April 1985. To fill in some of the missing values, we used the daily means of all available monitoring data as the basis for constructing our measures of PM_{10}. With all available data, there are observations for 75% of the days after 1 April 1985. Since many of the 20 monitoring stations were in operation for a short period, there is a maximum of 12 observations on any single day. Furthermore, the 6-day monitoring stations tend to operate on the same schedule, so many of the days have only the single daily monitor contributing to the daily mean.

In Cook County, PM_{10} levels are generally highest in the summer. Figure 1 shows the distribution of daily PM_{10} values by month. It is clear that mean levels are generally well below the EPA standard of 150 $\mu g/m^3$. In Table 2, the daily means from all available stations are compared with the values from the single daily monitoring station. These show close agreement, with three observations over the EPA standard for the daily station and two observations over 150 for the daily means.

In Salt Lake County, there were six PM_{10} monitors operating between June 1985 and December 1990. The monitoring network includes two daily stations. We use the observations from just one of the daily stations, station 12, in this analysis. Station 12 is centrally located in Salt Lake County. The second daily monitor is located in a more remote section of the county and was considered to be unreliable to use in measuring general exposure levels. Figure 1 shows the distribution of daily PM_{10} values by month for the centrally located daily station (station 12). The distribution of PM_{10} in Salt Lake County differs slightly from the distribution in Cook County. The overall levels are similar, though there are more days in Salt Lake County with PM_{10} levels over 150 $\mu g/m^3$. Unlike Cook County, there is an increase in overall levels in winter (December–February), though isolated occurrences of high particulate levels occur throughout the spring and summer. In Table 2, we present some summary statistics from the single daily station used in this analysis.

Meteorological data. The meteorological data used in this study are based on hourly surface observations taken at O'Hare International Airport (Cook County) and Salt Lake City International Airport (Salt Lake County). We extracted the data from the National Climatic Data Center's National Solar and Meteorological Surface Observation Network (1961–1990) database, which contains hourly surface observations in addition to solar radiation data. Our primary analyses concentrated on three meteorological variables: temperature, specific humidity, and barometric pressure. We excluded other variables such as solar radiation, cloud cover, wind speed, and wind direction. These variables were omitted to make our primary analyses more directly comparable with other research and because factors like wind may have more direct connection with PM_{10} than those included. For each variable we did include, we calculated the daily mean, based on hourly values. And, because weather may have a lagged effect on mortality, we also included the values of temperature, humidity, and pressure

Table 1. Mean daily mortality for nonaccidental causes of death

	Cook County					Salt Lake County	
	Elderly[a]	Total[b]	Circulatory[c]	Cancer	Respiratory	Elderly	Total
Winter	90.4	126.7	62.5	28.9	11.8	7.4	10.2
Spring	82.3	116.7	56.3	28.3	10.2	6.8	9.2
Summer	77.0	110.6	52.6	28.4	8.8	6.3	8.5
Fall	81.5	115.6	54.9	29.2	9.6	6.6	8.9

[a]Elderly mortality indicates the subset of these deaths among county residents aged 65 and older.
[b]Total mortality indicates the mean number of daily deaths of county residents of all ages, excluding accidental deaths, homicides, and suicides.
[c]Circulatory, cancer, and respiratory deaths are classified by the primary cause of death code listed on residents' death certificates.

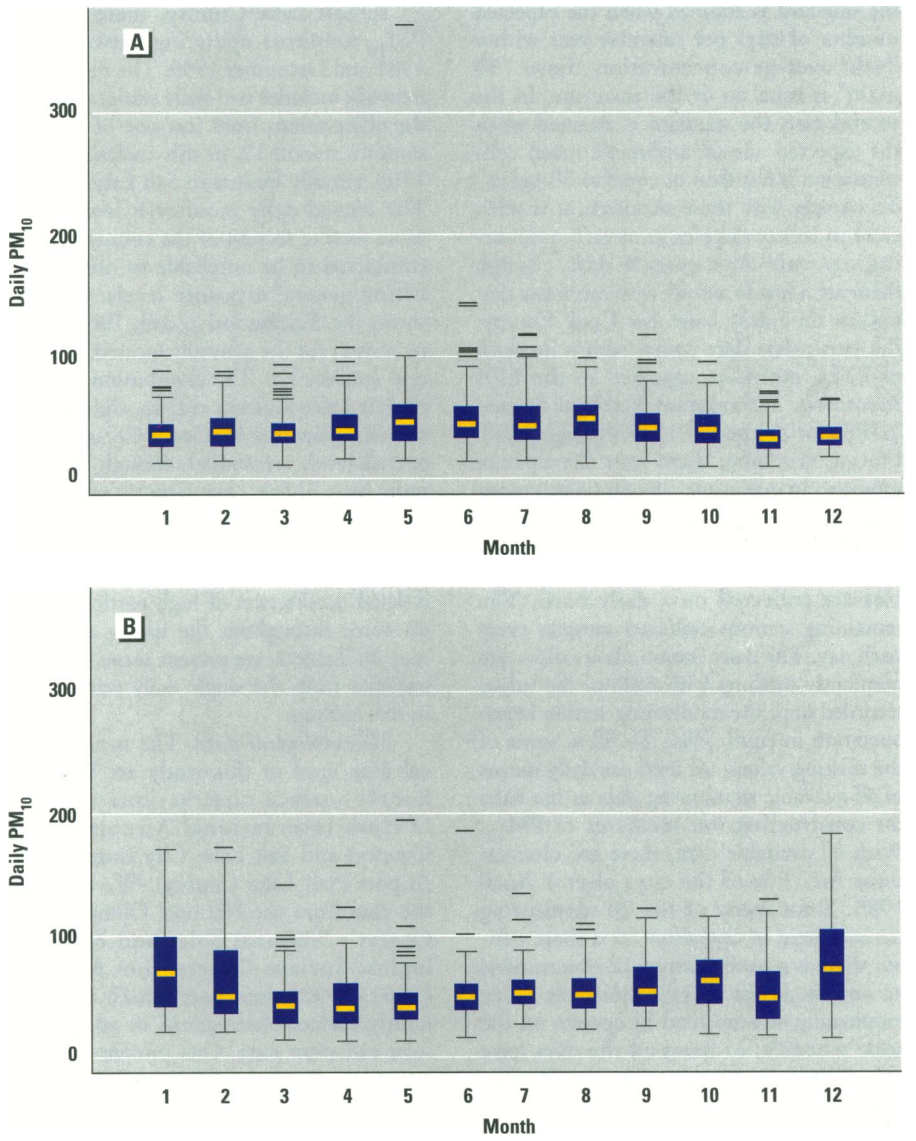

Figure 1. Daily particulate matter ≤10 µm (PM_{10}) by month for (A) Cook County and (B) Salt Lake County. Box plots by month showing the distribution of the daily network averages of PM_{10} observations for Cook County, and the observed values from the centrally located daily station in Salt Lake County.

Table 2. Summary of particulate matter ≤10 µm (PM_{10}) values[a]

	Min	25th Percentile	Median	75th Percentile	Max	No. days over 150 µg/m^3
Cook County						
Daily station	3	28	38	51	365	3
Network mean	4	27	37	50	365	2
Salt Lake County						
Daily station no. 12	9	33	48	67	194	13

[a]Statistics listing the minimum, 25th, median, 75th, and maximum PM_{10} values for daily stations in Cook and Salt Lake Counties, and the corresponding network averages for all available monitoring data.
[b]The number of days with PM_{10} values >150 µg/m^3 is listed in the final column.

from the 2 previous days. In other analyses, we considered the effect of wind chill in the winter and solar radiation and a heat index in the summer. These variables did not improve the prediction of mortality; the analyses are not included. The inclusion of wind speed and lagged wind speed in Cook County did not change the results from any of the models fit without wind.

Table 3 presents a summary of the meteorological data considered in various analyses. The data set containing the original hourly observations for these variables had only a few (nonsequential) missing values. We filled in the missing hourly observations by assigning the value from the previous hour, and then computed the daily mean values based on 24 observations.

Pollen data. Pollen data were obtained from the pulmonary unit at Grant Hospital, Chicago, Illinois, courtesy of Judith Young. During the study period, pollen counts were recorded on a daily basis, except for weekends and holidays, when cumulative samples were taken. To fill in daily pollen values from the cumulative values, we used a model to predict daily pollen from local meteorological conditions and then distributed the total pollen amounts to the individual days based on this model. We considered pollen from trees, mold spores, and ragweed.

Model Formulation

Our primary analyses modeled daily death counts as a Poisson process. For most analyses, we split the data by 3-month seasons and fit separate models within each season. Winter is taken as December–February; spring as March–May, etc. All season-by-season models include a yearly factor and a within-season trend (day) component. The specification of the trend component differs by season. For each season, we considered either a polynomial or a piecewise linear trend component and selected the shape that fit the data best. Although the covariates differ for different analyses, the basic model assumes that the daily death counts (Y) are Poisson-distributed with

$$\log(EY) = X\beta$$

where X contains terms corresponding to a yearly factor, a within-seasonal trend component, relevant meteorological covariates, and a measure of particulates. The parameters of the model were fit by the iterative, reweighted least-squares algorithm in the statistical software package Splus (MathSoft, Inc., Seattle, Washington) (*9*).

To account for a possible lagged effect of PM_{10}, we focused primarily on the 3-day PM_{10}, the average of the current day's PM_{10} together with the values for the 2 preceding days. Missing values were ignored, so the mean values were based on any available observations. We compared the results from these models with models that incorporated each of the 3 single-day values. We also did analyses using only the current day, a 2-day PM_{10} (today and yesterday), and a 5-day PM_{10} (today and 4 previous days). In essence, the results using the 3-day PM_{10} are consistent with these other choices of PM_{10} measures, so we only report a typical result from Cook County using the 5-day PM_{10} in the fall.

Auxiliary to the Poisson regression models used is a semi-parametric model which, through its nonparametric character, avoids the necessity of specification of special forms while allowing a reasonably

Table 3. Description of meteorological variables

Variable	Description
t_{mean}	Average daily temperature (°C) from hourly observations
*t*lag-1	Average temperature from 1 day before
*t*lag-2	Average temperature from 2 days before
q_{mean}	Average daily specific humidity (g/kg) from hourly observations
*q*lag-1	Average specific humidity from 1 day before
*q*lag-2	Average specific humidity from 2 days before
pr	Average daily station pressure (millibars) from hourly observations
*pr*lag-1	Average station pressure from 1 day before
*pr*lag-2	Average station pressure from 2 days before

Table 4. Candidate covariates for Poisson regression analyses based on results from semi-parametric modeling on elderly mortality

Month[a]	Cook County	Salt Lake County
January	day[b]	*pr*[c],t_{mean},day
February	q_{mean},*q*lag-1,*pr*,*pr*lag-2	day,*pr*lag-1,*pr*lag-2
March	day, 3-day PM_{10}[d]	
April		
May	t_{mean},*q*lag-2, 3-day PM_{10}	
June	*pr*	day, 3-day PM_{10}
July	t_{mean}	*q*lag-1,*pr*lag-2, 3-day PM_{10}
August	*t*lag-1,*pr*,*q*lag-1	
September	*q*lag-2,*pr*,*pr*lag-2, 3-day PM_{10}	day,t_{mean}
October	*pr*	
November	*q*lag-2	
December	day,*q*lag-1,*pr*lag-1	day,*t*lag-1

PM_{10}, particulate matter ≤10 µm.
[a]Active variables appearing in the month-by-month analyses using the semi-parametric model described in Appendix A.
[b]The variable day is the day of month (1–31).
[c]The meteorological variables are described in Table 3.
[d]The 3-day PM_{10} is the simple average of the observed network daily means for the concurrent day and 2 previous days.

accurate selection of important covariates. The details of the model as it was used are given in Appendix A. This model is used in several ways. Primarily, it was used to select relevant meteorological covariates and to focus on potentially important interactions as well as nonlinear functional forms for some of the covariates. Models selected in this fashion tend to be more parsimonious than models selected with standard stepwise procedures, with no loss of explanatory power. In addition, a month-by-month analysis using the semi-parametric model revealed that PM_{10} was usually an inactive factor.

By focusing on the months where PM_{10} does appear active, a possible connection with pollen was suggested. Accordingly, we obtained pollen data from the City of Chicago and introduced it in the month-by-month analyses of May and September, as well as in additional analyses covering August 15 to September 15, the ragweed season. In no case did any pollen variables appear as active factors in the semi-parametric model. Given the available pollen monitoring data, the observed PM_{10} effect in May and September is not explained by the presence of pollen particles.

With the focus on 3-day PM_{10}, the meteorological covariates that were considered at the first stage include the current day's values as well as values for the preceding 2 days. The particular covariates included for a season's analysis incorporated those found in the monthly analyses by the semi-parametric model. Table 4 shows the set of active factors for each month in both Cook County and Salt Lake County in the semi-parametric model. We considered each of these covariates as the candidate variables for inclusion in the Poisson regression models, along with the functional forms and interactions suggested by the fitted response surfaces from the semi-parametric model. To illustrate this use of the semi-parametric model, we include some plots of estimated effects of 3-day PM_{10}, temperature, pressure, and day-of-year for some selected months (Fig. 2). These effects are computed by conditioning the remaining variables on their median values, that is, by fixing them equal to their median values. The plots show the so-called Christmas effect on mortality, with a spike in the number of deaths around the beginning of January, the linear effect of PM_{10} in May and September, and the nonlinear effects of temperature and pressure. Using the combined list of covariates from the months composing each season, we used a stepwise variable selection technique to obtain a model without any measure of PM_{10}. Typically, this led to two or three meteorological covariates selected for each season to predict daily mortality. As a final step, we included the measure of PM_{10} and examined the direction and size of the corresponding coefficient.

To illustrate the importance of considering a season-by-season analysis, we also present results from an analysis combining the full year of observations for both Cook County and Salt Lake County. In this analysis, we fitted a yearly factor, a cubic time trend for each season, the meteorological covariates that were significant predictors of mortality in the season-by-season models, and seasonal interaction terms for selected meteorological covariates. We then compared the estimation of the PM_{10} effect from the models with and without PM_{10}-by-season interaction terms.

Results and Discussion

Empirical Evaluation in Cook County

There are several sets of results for Cook County. We first present full-year and season-by-season analyses using the Poisson regression model estimating daily death counts for individuals 65 and older (elderly mortality). Because daily death counts are high here, an ordinary (normal) regression model will give similar results. The linear predictors are detailed in Tables 5 and 6. As discussed in the previous section, the covariates other than the yearly factor and the PM_{10} variable were chosen using stepwise selection techniques based on the list of candidate covariates in Table 4. Other models and results for Cook County are summarized in Table 7.

In our full year analysis of Cook County, we conclude that it is necessary to estimate a separate PM_{10} effect for each season. Since the effect of meteorology differs by season (for example, increasing temperature acts as a stress factor in summer but decreasing temperature creates stress in winter), we began by considering models for the full year, which permitted separate estimates of the effect of weather within each season. Our final full-year model to predict elderly mortality from meteorology

includes separate seasonal terms for the yearly factors, the day-of-year effect, and temperature lagged 1 day. This permits the estimation of separate coefficients within each season for these terms. Other covariates whose effects do not vary significantly by season for Cook County include specific humidity for the concurrent day, 2-day lagged specific humidity, and station pressure for the concurrent day and previous day. We added the 3-day mean PM_{10} variable and compared the results from fitting a single estimate for the entire year with fitting separate estimates by season. The estimate for the single PM_{10} effect is 0.00054 with a standard error of 0.00020. Hence, an increase of 10 $\mu g/m^3$ of PM_{10} corresponds to approximately 0.54% more deaths, given constant levels of all other covariates. When the season-by-PM_{10} interaction term is added, the PM_{10} effect remains significant only in the spring and fall (Table 5). The estimated effects for the winter and summer are essentially zero. The chi-square test for the difference in deviance caused by inclusion of separate seasonal estimates for PM_{10} supports this inclusion with a p-value of approximately 0.001. To compare the overall effect of PM_{10} from this model, we calculated the predicted increase in the number of deaths in each season if PM_{10} were increased by 10 units. Specifically, we added 10 units to each of the observed values of PM_{10} and calculated the total number of predicted deaths. The overall predicted increase in mortality is 0.63%. A similar calculation, based on independent analyses of each month using the semi-parametric model, produces a 0.41% increase.

A finer tuned season-by-season analysis is obtained by fitting a separate model for each season. Here, we used the variables suggested by the semi-parametric models for the corresponding months to choose a parsimonious model predicting mortality from meteorology. The results for the separate seasonal analyses are presented in Table 6. The covariates included in the seasonal models vary significantly between seasons, suggesting that a separate model for each season may be more realistic than one full-year model. The PM_{10} coefficients and standard errors, however, are similar to the full-year analysis with the season-by-PM_{10} interaction terms. There is a significant effect in spring and fall, and no significant effect in the winter and summer.

The reported standard errors are calculated assuming independent observations. To check this assumption, we examined the autocorrelation structure of the standardized residuals for the full-year analysis. We computed the first seven lagged autocorrelations and found no correlations greater than 0.03. These values are all less than the approximate critical value of $2/(N)^{1/2}$ = 0.045. Furthermore, the autocorrelations were neither persistently positive nor negative. We conclude that there is no evidence of significant serial correlation. Other diagnostic plots of the residuals confirm that the modeling assumptions are reasonable.

To investigate the consistency of the PM_{10} effect for different populations, we modeled daily death counts from several subgroups within Cook County and for different measures of PM_{10}, like a 5-day mean instead of a 3-day mean. Because the largest estimated PM_{10} effect for elderly mortality is in the fall, we restricted attention to this season. These analyses included total mortality (nonaccidental deaths, all ages), elderly males and females, elderly blacks and non-blacks, and total mortality classified by disease categories, including circulatory disease, respiratory disease, and cancer. For each group, we refitted the semi-parametric model by month to obtain the list of candidate covariates for the Poisson regression analysis. Table 7 shows the results from the final models selected.

To address concern over potential weekday versus weekend effects in both PM_{10} and mortality, we refitted the model for elderly mortality in the fall season, detailed in Table 6, to subsets of the data determined by day of week. We first extracted observations falling on Wednesdays, Thursdays, and Fridays, because the 3-day PM_{10} variable for these days is unaffected by the decline in PM_{10} over the weekend. The resulting 3-day PM_{10} coefficient is given in Table 7; it is approximately one-half of the size of the coefficient when all the data are used. We

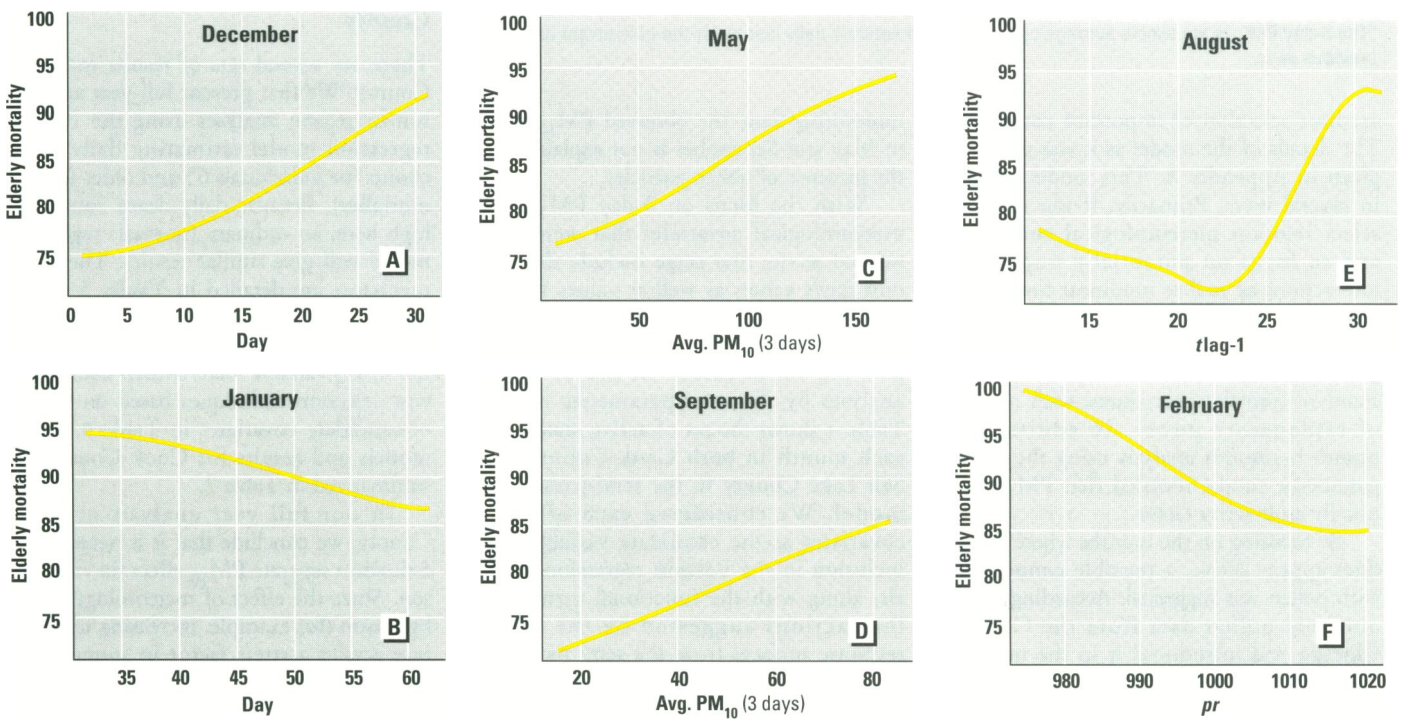

Figure 2. Some of the estimated effects for Cook County from the semi-parametric model. Predictions of elderly mortality holding all other variables constant at their median levels. (A,B) The day-of-month effect for December and January, highlighting the peak in the number of deaths around January 1; (C,D) the relationship between PM_{10} and mortality; (E,F) the potentially nonlinear dependence of mortality on meteorology. PM_{10}, particulate matter ≤10 µm; *t*lag-1, average temperature from 1 day before; *pr*, average daily station pressure from hourly observations.

Table 5. Full-year Poisson regression models: elderly mortality

Linear predictor log(*EY*)[a]		PM_{10} coefficient[b] (SE)
Cook County		
season * [year+poly(day,3)+poly(*t*lag-1,2)]+*q*+*q*lag-2+*pr*+poly(*pr*lag-1,2)+3-day PM_{10}		0.00054 (0.00020)
season * [year+poly(day,3)+poly(*t*lag-1,2)+3-day PM_{10}]+*q*+*q*lag-2+*pr*+poly(*pr*lag-1,2)	Winter	-0.00001 (0.00047)
	Spring	0.00083 (0.00034)
	Summer	-0.00028 (0.00036)
	Fall	0.00195 (0.00047)
Salt Lake County		
Model 1		
season * [year+poly(day,3)+*pr*+poly(*q*lag-1,2)]+t_{mean}+3-day PM_{10}		-0.00025 (0.00043)
Model 2		
season * [year+poly(day,3)+poly (*q*lag-1,2)]+t_{mean}+3-day PM_{10}		0.00008 (0.00041)

PM_{10}, particulate matter ≤10 µm.
[a]The left-hand side shows the models fit to predict daily mortality, where brackets indicate the interaction terms included and poly(variable, *n*) indicates a polynomial term for the given variable of order *n*. Specification of an interaction implies inclusion of all lower-order terms.
[b]The right-hand side shows the estimated effects of the PM_{10} variable, along with estimated standard errors.

Table 6. Seasonal Poisson regression models: elderly mortality

Linear predictor log(*EY*)[a]		PM_{10} coefficient[b] (SE)
Cook County		
Winter	year+day+Jan+Feb+*pr*+poly(*pr*lag-1,2)+3-day PM_{10}	0.00024 (0.00046)
Spring	year+*q*lag-2+3-day PM_{10}	0.00088 (0.00030)
Summer	year+poly(day,2)+poly(*t*lag-1,2)+*pr*+*pr*lag+3-day PM_{10}	-0.00024 (0.00035)
Fall	year+*q*lag-2+3-day PM_{10}	0.00138 (0.00040)
Salt Lake County		
With station pressure		
Winter	year+day+Jan+Feb+*t*lag-1+*pr*+3-day PM_{10}	0.00021 (0.00057)
Spring	year+3-day PM_{10}	0.00032 (0.00091)
Summer	year+poly(day,2)+3-day PM_{10}	-0.00027 (0.00124)
Fall	year+day+Oct+*pr*+Nov+3-day PM_{10}	-0.00131 (0.00094)
Without station pressure		
Winter	year+day+Jan+Feb+*t*lag-1+3-day PM_{10}	0.00056 (0.00054)
Fall	year+day+Oct+Nov+3-day PM_{10}	-0.00113 (0.00093)

PM_{10}, particulate matter ≤10 µm.
[a]Models for mortality estimated separately within each season are listed on the left-hand side.
[b]Estimated coefficients and standard errors are shown on the right.

Table 7. Summary of regression models, Cook County, fall season

Population[a]	Linear predictor log(*EY*)[b]	PM_{10} coefficient[c] (SE)
Total mortality	year+t_{mean}+*q*lag-2+poly(*pr*lag-2,3)+3-day PM_{10}	0.00080 (0.00040)
Males 65+	year+poly(*q*lag-2,2)+t_{mean}+3-day PM_{10}	0.00159 (0.00069)
Females 65+	year+*q*lag-2+3-day PM_{10}	0.00087 (0.00054)
Blacks 65+	year+poly(*q*lag-2,3)+3-day PM_{10}	0.00166 (0.00089)
Whites, others 65+	year+*q*lag2+3-day PM_{10}	0.00134 (0.00045)
Circulatory deaths	year+poly(*q*lag-2,2)+poly(*pr*lag-2,3)+3-day PM_{10}	0.00064 (0.00052)
Respiratory deaths	year[*t*lag-2+poly(q_{mean},2)]+3-day PM_{10}	0.00220 (0.00125)
Cancer deaths	year+poly(q_{mean},3)+poly(*t*lag-2,2)+3-day PM_{10}	0.00162 (0.00071)
Elderly mortality	year+poly(*q*lag-2,3)+5-day PM_{10}	0.00158 (0.00047)
Wed, Thurs, Fri	year+*q*lag-2+3-day PM_{10}	0.00075 (0.00061)

PM_{10}, particulate matter ≤10 µm.
[a]The population subgroups for each analysis are listed in the left-hand column.
[b]The final models are indicated in the middle column.
[c]The corresponding coefficients and standard errors for the PM_{10} variable are listed in the right-hand column.

also analyzed each day of the week individually. Although all of the 3-day PM_{10} coefficients were positive, only the coefficient based on the Sunday data was significantly different from zero. The average of the seven daily coefficients is 0.00135, comparable to the coefficient of 0.00138 obtained in our original Poisson regression analysis of elderly mortality for fall (Table 6). Similar effects were observed in the spring. We interpret these results as inconclusive, neither supporting nor denying a weekday effect.

Although there appear to be inconsistencies in Table 7 (for example, a significant effect of PM_{10} on males but not on females), the difference of the two effects may be insignificant. In our analyses, the coefficient for cancer deaths is greater than the coefficient for circulatory deaths. This ordering is reversed from the numbers reported for Philadelphia (*1*) but, again, the differences in the coefficients may not be significantly different from zero. The lack of significance for blacks is due to the greater standard error resulting from the smaller size of the black population in Cook County. The estimated coefficient for elderly blacks is actually larger than the estimated coefficient for the whites and others category. The distinction between using the 5-day PM_{10} rather than the 3-day PM_{10} is to reduce the size of the effect somewhat, from 0.00195 to 0.00158, but it remains significant.

Empirical Evaluation in Salt Lake County

The analyses for Salt Lake County were carried out in similar fashion to those carried out in Cook County. The semi-parametric model was used on transformed (square-root of) mortality to ameliorate the effect of non-normality and nonconstant variances in the presence of small counts. The analyses proceeded as before from the variables in Table 4 to the models in Table 6.

The semi-parametric model identified PM_{10} as active in June and July. An estimated effect plot for July indicated that the effect of PM_{10} in July was oscillatory (as in March in Cook County) rather than monotone as in June (or as in May and September in Cook County; see Fig. 3). The Poisson regression analysis, however, did not support evidence of a PM_{10} effect in the summer. In fact, for the full-year and seasonal models, PM_{10} was never a significant predictor of elderly mortality in Salt Lake County.

For the full-year analysis, the single estimate of the PM_{10} effect is -0.00025 with a standard error of 0.00043 (Table 5). The full-year model including the season-by-PM_{10} interaction term fails to indicate a significant PM_{10} effect in any single season. Furthermore, unlike Cook County, the chi-square test for the difference in residual deviance does not support the inclusion of a season-by-PM_{10} interaction term. To investigate whether a possible PM_{10} effect is being masked by the presence of station pressure in this model, we refit the full-year model without station pressure as one of the candidate covariates. The selected model is identical, except for the deletion of station pressure. In this model (Table 5), there is also no significant PM_{10} effect. Additionally, the interaction term between season and PM_{10} fails to indicate a significant PM_{10} effect within any single season.

We also fitted separate models for each

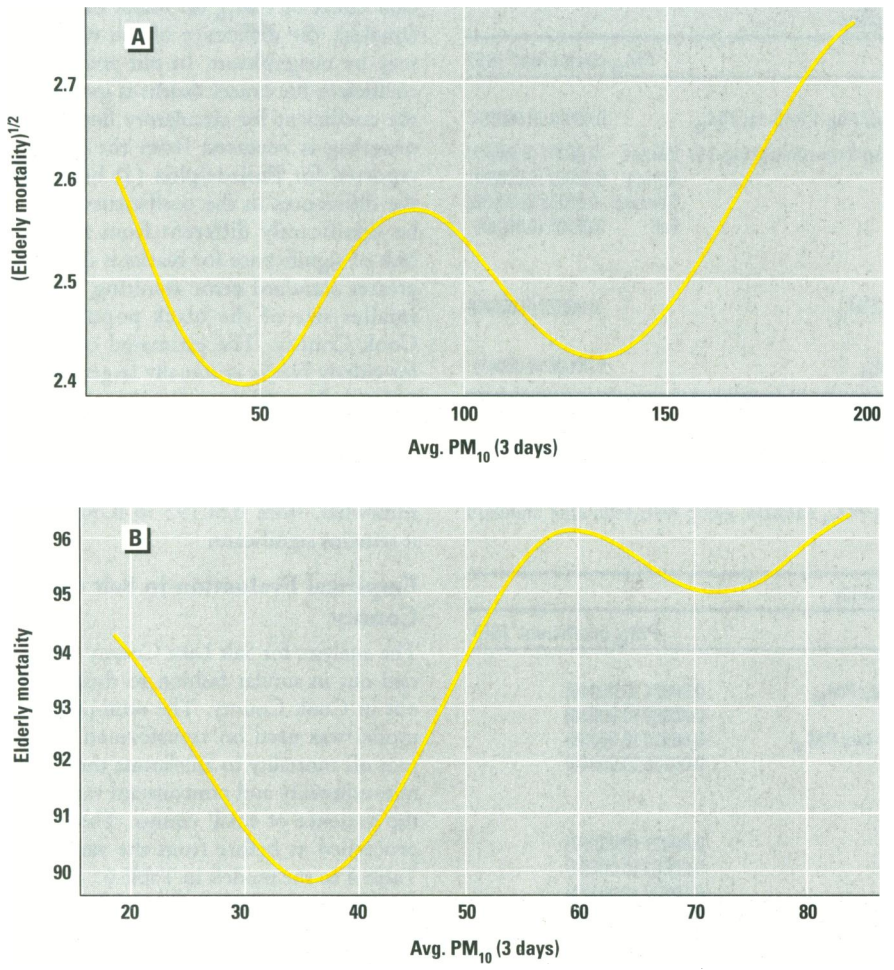

Figure 3. Estimated oscillatory effects from the semi-parametric model. In some months, particulate matter ≤10 µm (PM_{10}) appears as an active factor in the semi-parametric model, but the effect appears to be spurious.

season, as reported in Table 6. Here, we present results both with and without station pressure. Regardless of the inclusion of station pressure, PM_{10} never shows up as a significant predictor of mortality.

Summary

In summary, we analyzed data from Cook County, Illinois, and Salt Lake County, Utah, to assess the connections among mortality, particulates (PM_{10}), and weather. We found that season plays a strong role in Cook County. We found inconsistent results: no effect of PM_{10} was found in Salt Lake County in any season; no effect was found in Cook County in winter and summer; small, positive PM_{10} effects were found in Cook County in the spring and fall, and, more specifically, in the months of May and September.

One of the reasons for using multiple regression techniques is to remove the possible confounding effects of weather and possibly other pollutants. We demonstrate in Appendix B that weather conditions and airborne particulates are indeed associated in both Cook County and Salt Lake County. It is also generally accepted that weather conditions affect mortality rates. Under these circumstances, it is difficult to rule out the possibility that there is some common third cause of both elevated particulate levels and mortality. Perhaps more importantly, it makes it very difficult to understand the impact of having potentially large errors in the explanatory variables. Outdoor monitors, as well as airport weather data, are crude approximations of individual exposure levels. And any effort to include additional pollutants, like ozone, which is highly correlated with both particulates and weather, can also produce confusing results in the multiple regression setting. While we have not addressed all of these issues in detail, we have attempted to highlight some of the limitations of regression analysis in the discussion of our results.

We intentionally selected two counties with very different characteristics. Although our results were quite different depending on the location, we do not know whether this is due to differences in the populations, differences in the composition of PM_{10}, differences in weather that were not adequately modeled, or some other variable. To deduce causal relationships from the type of PM_{10} and mortality data available requires these kinds of considerations, which go beyond regression analysis. We have provided additional evidence with which to judge whether the association between particulate levels and mortality is consistent across settings. Our results do not show as much consistency as previously published analyses.

Appendix A. Semi-parametric Model

On day i in year j, with meteorological condition *met* and PM_{10} value *pm*, where *met* is a nine-dimensional vector of the meteorological variables listed in Table 3 and *pm* could be any of the PM_{10} measures used in the analyses, let $\mathbf{x} = (pm, met, i, j)$. The vector $\mathbf{x} = (\zeta_1, \ldots, \zeta_{12})$ is 12-dimensional. The response $y(\mathbf{x})$ (mortality) is assumed to be a realization of a stochastic process, $Y(\mathbf{x})$:

$$Y(\mathbf{x}) = \beta_j + Z(\mathbf{x}) + \varepsilon_{ij} ,$$

where β_j are constants, $j = 1,2, \ldots,6$, $Z(\mathbf{x})$ is a zero mean Gaussian process with covariance function $\mathrm{cov}[Z(\mathbf{x}), Z(\mathbf{x}')] = \sigma_Z^2 R(\mathbf{x},\mathbf{x}')$ to be specified later, and $\varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2 I)$. For more discussion on the use of this technique for modeling response surfaces, see Sacks et al. (*10*), and the references cited therein.

Assume, as in Sacks et al. (*10*), that the covariance between $Z(\mathbf{x})$ and $Z(\mathbf{x}')$ is

$$\sigma_Z^2 R(\mathbf{x},\mathbf{x}') = \sigma_Z^2 \exp\left(-\sum_{k=1}^{12} \theta_k \left|\zeta_k - \zeta_k'\right|^{P_k} \right)$$

where $\mathbf{x} = (\zeta_1, \ldots, \zeta_{12})$, $\mathbf{x}' = (\zeta_1', \ldots, \zeta_{12}')$, $\theta_k \geq 0$; $k = 1, \ldots, 11$, $\theta_{12} = 0$ and $1 \leq p_k \leq 2$; $k = 1, \ldots, 12$; θ_{12} corresponds to the year variable. This class of stationary processes provides us with a wide range of functions.

Given the data $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$ for q consecutive years starting from year 1 (1985) with n_j data points in year j and $n_1 + \ldots + n_q = n$ and, provided σ_Z, σ_ε, and $R(\cdot,\cdot)$ are known, the best linear unbiased predictor $\hat{y}(\mathbf{x})$ at a new point $\mathbf{x}$ in year j can be written as

$$\hat{y}(\mathbf{x}) = \hat{\beta}_j + \hat{Z}(\mathbf{x}) = \hat{\beta}_j + \mathbf{r}'(\mathbf{x}) C^{-1}\left(\mathbf{y} - F\hat{\boldsymbol{\beta}}\right),$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, $C = \mathrm{Corr}(y) = (\sigma_Z^2/\sigma^2) R + (\sigma_\varepsilon^2/\sigma^2) I$, where $\sigma^2 = \sigma_Z^2 + \sigma_\varepsilon^2$, and $R = \{R(\mathbf{x}_i, \mathbf{x}_j), 1 \leq i \leq n; 1 \leq j \leq n\}$, the $n \times n$ matrix of correlations among values of Z at the data points, $\mathbf{r}(\mathbf{x}) = (\sigma_Z^2/\sigma^2) [R(\mathbf{x}_1,\mathbf{x}), \ldots, R(\mathbf{x}_n,\mathbf{x})]'$,

$$F = \begin{pmatrix} \vec{1}_{n_1 \times 1} & \vec{0} & \cdots & \vec{0} \\ \vec{0} & \vec{1}_{n_2 \times 1} & \cdots & \vec{0} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{0} & \vec{0} & \cdots & \vec{1}_{n_q \times 1} \end{pmatrix}_{n \times q},$$

and $\hat{\boldsymbol{\beta}}_1 = (\hat{\beta}_1, \ldots, \hat{\beta}_q)' = (F'C^{-1}F)^{-1} F'C^{-1}\mathbf{y}$, which is the usual generalized least-squares estimate of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)'$.

The parameters σ_Z, σ_ε, θ, and p are fit by maximum likelihood. Cross-validation is used to assess variability of estimates. Values of p indicate smoothness of the response surface as a function of the corresponding variables. Larger values of θ usually indicate greater importance of the corresponding variables if the variables are on normalized scales. During the covariate selection procedure, those coefficients (θ) which are zero are the factors not included; the others are selected.

Appendix B. The Problem of Confounding

To examine the confounding relationship between PM_{10} and the meteorological variables, a forward-selection ordinary least-squares regression analysis was performed with log PM_{10} (the natural logarithm of today's PM_{10}) serving as the response variable and the meteorological variables serving as the covariates. The meteorological variables in the PM_{10} analysis were those included in the mortality analysis. The same seasonal structure was maintained for the PM_{10} analysis as for the mortality analyses.

Cook County. As mentioned earlier, PM_{10} levels were highest in the spring and summer while fall and winter levels were depressed. The R^2 values from the final models based on the forward-selection ordinary least-squares regression analyses ranged from a low of 20% in the winter to a high of 50% in the summer. Thus the relationship was strongest during the season with the highest PM_{10} levels. With the exception of the 2-day lag temperature term ($t$lag-2) in the fall, the regression coefficients for the various temperature terms were positive. Today's temperature (t_{mean}) showed up in all seasons with the exception of summer, while the square of today's temperature showed up in all seasons. All seasons except winter exhibited a strong rise in PM_{10} with increasing temperature. The coefficients on the specific humidity terms were negative. Yesterday's specific humidity ($q$lag-1) was important in all seasons, while today's specific humidity (q_{mean}) showed up in spring and fall. A quadratic term [($q$lag-2)^2] showed up in the summer. These main-effect results are consistent in the sense that warmer, drier conditions contribute to increased levels of particulate matter. Interaction plots generally indicated that at low temperatures PM_{10} levels increased with increasing specific humidity, while the reverse was true at higher temperatures. Station pressure (2-day lagged variable, prlag-2) showed up only in the fall and then was positive

Salt Lake County. The amount of variation in PM_{10} explained by the meteorological covariates ranged from 41% in fall to a high of 53% in winter (a time of high PM_{10} levels). In contrast to Cook County, station pressure was a significant variable in all seasons in addition to temperature and specific humidity. Station pressure lagged 1 day (prlag-1) was the first variable to enter the forward selection process in fall and winter, where it added 25% and 42% to the R^2 value, respectively. The sign of the regression coefficient on the pressure terms was positive for all seasons. This strong association between pressure and particulate levels during fall and winter may have resulted from the occurrence of capping inversions which are associated with synoptic-scale high pressure systems. Given the nature of the landscape, these inversions would tend to trap pollutants near the earth's surface. In spring and summer, temperature terms were the first to enter the forward-selection process. The signs on the temperature terms varied with the season and within the season for different terms. Specific humidity terms entered all seasons in a negative manner except for winter. In spring and summer, PM_{10} levels generally increased as temperature increased; in winter PM_{10} levels decreased as temperatures rose. In fall an initial decrease in PM_{10} levels as temperatures rose turned to an increase in PM_{10} levels as temperatures moved above 7°C. In winter, summer, and fall PM_{10} levels initially increased with rising humidity levels and then began to drop as humidity continued to rise. In spring PM_{10} levels decreased as humidity increased. Results on fitting mortality to weather variables alone, without PM_{10}, indicated that temperature, humidity and pressure are all implicated (Tables 5 and 6).

References

1. Schwartz J, Dockery DW. Increased mortality in Philadelphia associated with daily air pollution concentrations. Am Rev Respir Dis 145:600–604 (1992).
2. Pope CA III, Schwartz J, Ransom MR. Daily mortality and PM_{10} pollution in Utah Valley. Arch Environ Health 47:211–217 (1992).
3. Slipp D. Bad things come in small particles. Wall Street Journal, 24 April 1991;B1.
4. Hilts PJ. Studies say soot kills up to 60,000 in U.S. each year. The New York Times, 19 July 1993;A1.
5. Brown W. Dying from too much dust. New Sci (March 12):12–13 (1994).
6. U.S. EPA. National primary and secondary ambient air quality standards. Fed Reg 36:22384 (1971).
7. Mosteller F, Tukey JW. Data analysis and regression: a second course in statistics. Reading, MA:Addison-Wesley, 1977.
8. Fairley D. The relationship of daily mortality to suspended particulates in Santa Clara County, 1980–1986. Environ Health Perspect 89:159–168 (1990).
9. Chambers J, Hastie T. Statistical models in S. New York:Chapman and Hall, 1993.
10. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. Stat Sci 4:409–435 (1989).

International Symposium

Evaluation of Butadiene & Isoprene Health Risks

27–29 June 1995
Inn at Semi-ah-Moo Blaine, Washington

A Three Day Symposium Preceding ICT-VII On Butadiene and Isoprene Issues
Metabolism and Pharmocokinetics Biological Effects Epidemiology Assessment Perspectives and methods

Jointly organized by AIHA▪API▪CEFIC▪CIIT▪CMA▪EPA▪ECETOC▪HEI▪IARC▪IISRP▪IPCS▪NIEHS▪SOT

INQUIRIES
John Ludwig, Symposium Coordinator
International Institute of Synthetic Rubber Producers, Inc.
2077 S. Gessner, Ste 133 Houston TX 77063
Tel: 713-783-7511 ext. 255 Fax: 713-783-7253