LBNL-43583
Preprint

ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY

An Exploratory Analysis of the Relationship between Mortality and the Chemical Composition of Airborne Particulate Matter

Feng C. Tsai, Michael G. Apte, and Joan M. Daisey

Environmental Energy Technologies Division

December 1999

Submitted to
Inhalation Toxicology

REFERENCE COPY
Does Not
Circulate
Library Annex Reference
Lawrence Berkeley National Laboratory

Copy 1

LBNL-43583

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBNL-43583

An Exploratory Analysis of the Relationship Between Mortality and the Chemical Composition of Airborne Particulate Matter

Feng C. Tsai,[1,2] Michael G. Apte,[1] and Joan M. Daisey[1]

Environmental Energy Technologies Division
[1]Indoor Environment Department
Ernest Orlando Lawrence Berkeley National Laboratory
Berkeley, California 94720

[2]School of Public Health
University of California
Berkeley, California 94720

December 1999

This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

ACKNOWLEDGEMENTS

The authors wish to thank Bill Fisk, Kirk R. Smith, William Nazaroff, Alan Hubbard, and Catherine Koshland for their careful reviews and comments. This research was supported by the Lawrence Berkeley National Laboratory Research & Development Fund, under Department of Energy Contract No. DE-AC03-76SF00098.

ABSTRACT

We explored relationships between daily mortality and the major sources of airborne particulate matter (PM) using a newly developed approach, Factor Analysis and Poisson Regression (FA/PR). We hypothesized that by adding information on PM chemical speciation and source apportionment to typical PM epidemiological analysis, we could identify PM sources that cause adverse health effects. The FA/PR method was applied to a merged dataset of mortality and extensive PM chemical speciation (including trace metals, sulfate and extractable organic matter) in New Jersey.

Statistically significant associations were found between mortality and several of the FA-derived PM sources, including oil burning, industry, sulfate aerosol, and motor vehicles. The FA/PR method provides new insight into potentially important PM sources related to mortality. For the dataset we analyzed, the use of FA/PR to integrate multiple chemical species into source-related PM exposure metrics was found to be a more sensitive tool than the traditional approach using PM mass alone.

LIST OF FIGURES

Figure 1.	ATEOS sampling sites	21
Figure 2.	Newark data	22
Figure 3.	Camden data	23
Figure 4.	Elizabeth data	24
Figure 5.	The FA/PR method	25

LIST OF TABLES

Table 1	Data summary for ATEOS sampling periods: arithmetic mean (standard deviation)	26
Table 2	Results of simple Poisson models: parameter estimate (standard error)	27
Table 3	Results of factor analysis in Newark (source tracers underlined)	28
Table 4	Results of FA/PR analyses: p values	29
Table 5	Relative risks associated with an increase of 10 $\mu g/m^3$ by the traditional approach using total PM mass and [FA/MR]/PR	30
Table 6	Socioeconomic status and population data in New Jersey	31

TABLE OF CONTENTS

Acknowledgements..... ii

Abstract..... iii

List of Figures..... iv

List of Tables..... v

Introduction..... 2

Data..... 4

Environmental data..... 4

Mortality data..... 5

Analysis methods..... 6

Simple Poisson model..... 6

Factor analysis and Poisson regression (FA/PR)..... 7

[Factor analysis/multiple regression]/ Poisson regression ([FA/MR]/PR)..... 10

Results..... 11

Simple Poisson model..... 11

FA/PR..... 12

[FA/MR]/PR..... 13

Model comparison..... 14

Discussion..... 15

References..... 32

Appendices..... 36

A. Results of basic Poisson models..... 36

B. Output of Factor Analysis..... 38

C. Results of FA/PR..... 41

D. Model comparison..... 44

INTRODUCTION

Prior epidemiological studies have provided compelling and relatively consistent evidence of associations between ambient particulate matter (PM) levels and adverse health effects, including increases in mortality, increases in respiratory symptoms, hospitalizations, and disease, along with declines in lung function (see review paper by Pope *et al.,* 1995). Unlike other criteria pollutants, PM has no specified chemical formula under the Clean Air Act. The National Ambient Air Quality Standard (NAAQS) for PM is set by particle size and mass concentration, but PM is made up of different chemical components and particle sizes emitted from various sources. Limited data are available on the day-to-day variations in the chemical constituents of PM. Thus, traditional PM epidemiological studies generally use ambient PM mass concentrations (e.g., Total Suspended Particle, PM_{15}, PM_{10}, or $PM_{2.5}$, or mass-equivalent, such as visibility, Coefficient of Haze, or British Smoke) or concentrations of a single PM component (such as SO_4^{2-}) as metrics of human exposure. These epidemiological analyses implicitly assume that the same amount of PM *mass*, regardless of its constituents and size distribution, produces equivalent adverse health effects in humans. In reality, PM mass contains a complex mixture of multiple chemical components across a wide distribution of sizes. PM mass may only serve as a surrogate for the specific constituents causing the adverse health effects. Possible causal agents for PM-related health effects may be particle mass concentration, particle number concentration, acid aerosols, or specific chemicals or combinations of chemicals composing the particles, but to date the biological mechanisms are not fully understood. If chemical differences are significant in the induced health effects, PM from different sources with different

chemical compositions should display different relationships for a given adverse health effect.

Factor Analysis (FA) has been used extensively to identify major PM sources. Poisson Regression is often used in epidemiological studies to characterize the relationship between mortality (rare events) and ambient air pollution. In this study, it was first proposed to integrate these two analytical techniques, Factor Analysis/ Poisson Regression (FA/PR), to understanding the relationship between mortality and individual PM sources. FA/PR was applied to a unique data set with mortality and extensive PM chemical speciation measurements (including trace metals, sulfate, and extractable organic matter). In addition to FA/PR, the relationship between mortality and single PM-mass related exposure metrics including IPM (inhalable particulate matter, $d_{50} \leq 15$ μm), FPM (fine particulate matter, $d_{50} \leq 2.5$ μm), SO_4^{2-}, and three fractions of extractable organic matter within PM was also investigated.

DATA

1. ENVIRONMENTAL DATA

The ambient pollution data were obtained from the Airborne Toxic Element and Organic Substances study (ATEOS, Lioy and Daisey, 1987). The ATEOS study was conducted between 1981 to 1983 at three New Jersey sites: Newark, Elizabeth, and Camden (map shown in Figure 1). The sampling site in Newark was in an industrial-residential area of the Ironbound district where heavy traffic and various industrial sources were present. In Camden, the sampling site was in a commercial-residential area close to refineries and industrial facilities (in Philadelphia). In Elizabeth, the sampling site was in a residential area close to a business district. The sampling sites of Newark and Elizabeth were about 8 km apart. The samplers were on the roof of the buildings or trailers. Environmental measurements taken were IPM, FPM, trace metals (Pb, Mn, Fe, Cd, V, Ni, Zn, Cu), and sulfate constituents of IPM collected by continuous 24-hour Hi-Volume samplers. Figures 2-4 show the daily variation in IPM and sulfate at the three sampling sites. Three fractions of extractable organic matter (EOM), including cyclohexane-solubles (CX), dichloromethane-solubles (DCM), and acetone-solubles (ACE), were also measured (Daisey, 1987). The sampling periods were two consecutive summers and winters with 39 sampling days in each period: July 6-August 14, 1981; January 18-February 25, 1982; July 6-August 14, 1982; and January 17-February 25, 1983. In addition to the pollutants measured by ATEOS, maximum 1-hour daily CO and daily temperature (maximum daily temperature and minimum daily temperature) were also included in the following analyses (data summarized in Table 1).

The ambient PM monitoring data obtained from the ATEOS study were used as exposure metrics for local residents in New Jersey to evaluate the relationship between mortality and particulate air pollution in the following analyses. Although the *absolute* concentrations of ambient PM and human exposure may be different, studies have shown that the daily changes of ambient PM correlate well with indoor PM or human exposure in a *relative* scale (Janssen, 1998). Using ambient data to represent human exposure is not perfect, but it is the best we can do with the data set for this air pollution epidemiological study.

2. MORTALITY DATA

New Jersey mortality data were obtained from Public Use Data Tape Files: Mortality Detail for 1981-1983 (U.S. Department of Health and Human Services, National Center for Health Statistics). These data files, recorded by year, contain information for every death of a U.S. resident during that year. A subset of the New Jersey mortality data for the 3 cities was extracted to match the ATEOS sampling sites by location of residence, not location of occurrence of death. Accidental and homicide deaths (ICD codes > 800) were excluded from our analyses. Two death categories were defined for statistical analyses in this study: total daily deaths (TDD) and cardiovascular and respiratory daily deaths (CRDD; ICD codes: 480-486, 490-496, and 390-448). Population number, and averaged TDD and CRDD in the three cities are summarized in Table 1.

ANALYSIS METHODS

We applied three approaches in the data analyses. First, as is typical in PM epidemiological studies, simple Poisson Regression (PR) was applied to investigate the association between TDD or CRDD and individual PM exposure metrics. In the second method, we performed FA on the chemical speciation data (excluding PM mass; see Table 1 for species included in FA) to resolve PM source-related factors and used these factors as exposure metrics in PR to identify statistically significant PM sources associated with mortality. Three variations of FA/PR analyses were compared to the simple Poisson model using a single exposure metric, PM mass. Finally, multiple regression (MR) analysis was used to estimate the risk associated with each PM source significantly associated with mortality identified by FA/PR.

SIMPLE POISSON MODEL-- USING A SINGLE PM EXPOSURE METRIC

The Poisson model is commonly used to model discrete events (such as mortality) that occur infrequently in time or space. It is sometimes called "distribution of rare events." The Poisson model is as follows:

$$\log [E(Y)]=\mathbf{X}\beta,$$

where E(Y) is the expected value of daily deaths, Y; **X** is the matrix of covariates; and β is the vector of estimated regression coefficients.

Poisson regression was applied to assess the association between single PM exposure metrics (IPM mass, FPM mass, sulfate, and three organic fractions-- CX, DCM and ACE) and daily mortality count (TDD and CRDD) in each city, adjusted for the daily average temperature (T_{avg}). Other possible confounders, time of week (weekday/weekend) and time of year (winter/summer), were evaluated by adding binary

variables ("WEEK"=0 for weekdays and "WEEK"=1 for weekend; "SEASON"=0 for winter and "SEASON"=1 for summer) in the Poisson models. These added binary variables were not found to be significant.

The Generalized Estimating Equation (GEE) (Liang *et al.*, 1986) with autoregressive correlation structure was used to account for the possible autocorrelation of the dependent variable, mortality. Lag effects, increased mortality as a result of previous cumulative PM exposure, were also evaluated in Poisson models with single exposure metrics. Mortality data were regressed against two types of lagged pollutant concentrations, single-day concentration (lag-1, lag-2 and lag-3), and the averages of current and preceding 1, 2, and 3 days concentrations.

FACTOR ANALYSIS AND POISSON REGRESSION (FA/PR) -- IDENTIFICATION OF PM SOURCES SIGNIFICANTLY ASSOCIATED WITH MORTALITY

Factor analysis (FA), one type of receptor modeling that requires no prior knowledge of source emission rate and meteorological conditions, has been used extensively to identify important sources of ambient pollution (e.g., Kleinman, 1977; Daisey *et al.*, 1981; Cox *et al.*, 1981; Hopke, 1985; Morandi, 1985; Morandi *et al.*, 1991). Receptor models, reviewed in Chapter 1, use the differences in physical or chemical properties of the emissions from different sources to apportion source contributions. FA converts multiple correlated environmental data into a reduced number of conceptually meaningful independent vectors, called factors. Factors are hypothetical variables selected to reproduce the correlation of measured chemical species using the minimum number of vectors. In factor analysis, a *common factor* is an unobservable, hypothetical variable that contributes to the variance of at least two of the observed variables; a *unique*

factor is an unobservable hypothetical variable that contributes to the variance of only one of the observed variables. The model for common factor analysis posits one unique factor for each observed variable. The equation for the common factor model is:

$$y_{ij} = x_{i1}b_{1j} + x_{i2}b_{2j} + x_{i3}b_{3j} + \ldots + x_{iq}b_{qj} + e_{ij}$$

where y_{ij} is the value of the ith observation on the jth variable.

x_{iq} is the value of the ith observation on the qth common factor.

b_{qj} is the regression coefficient of the qth common factor for predicting the jth variable.

e_{ij} is the value of the ith observation on the jth unique factor.

q is the number of common factors.

It is assumed for convenience that all variables have a mean of 0. In matrix terms, these equations reduce to

$$Y=XB+E$$

where X is the matrix of factor scores and B is the factor pattern. Two critical assumptions are made for the preceding equation: the unique factors are uncorrelated with each other and the unique factors are uncorrelated with the common factors (SAS, 1989).

Factor loading represents the correlation of the chemical species within each factor. Sources can be identified by comparing the chemical species that have high loadings in each factor with the known source signatures or tracers from PM emissions. Examples of source tracers for geological sources are Mn and Fe and for oil burning sources are V and Ni. Factor scores, the composite measures of factors resolved from FA, are PM source-specific transformations of the original measurements of chemical

species. These factor scores were used as exposure metrics to assess the relationship between daily mortality and individual PM sources (Figure 5).

FA was previously applied to the ATEOS data by Morandi *et al.* (1991) for IPM source apportionment. In Morandi's study, extreme outliers (about 12% of total samples) were excluded from the factor analysis because the purpose of those studies was to identify major PM emission sources in New Jersey. The objective of our study, however, was to assess the relationship between changes in ambient pollution levels and changes in daily mortality. Therefore, leaving out all high-pollution days (air pollution episodes during the ATEOS sampling period reported by Lioy *et al.*, 1985) as Morandi *et al.* did would not be appropriate because the mortality tends to increase during air pollution episodes (shown in Figures 2-4). Factor Analysis with Varimax rotation (SAS, 1989) was applied to the entire ATEOS dataset of trace elements, sulfate, and CO. Factor scores were then regressed against mortality in a Poisson model to identify the PM sources that were significantly associated with mortality.

The FA/PR method cannot be directly used to quantify the risk associated with each source because factor scores are normalized with mean zero and standard deviation equal to one. Therefore, we incorporated multiple regression into FA/PR to estimate the risk of significant sources.

[FACTOR ANALYSIS/MULTIPLE REGRESSION]/POISSON REGRESSION ([FA/MR]/PR): ESTIMATION OF RISKS ASSOCIATED WITH PM SOURCES

Factor Analysis is often used as an exploratory tool to identify source tracers or patterns in a qualitative manner. Quantitative source apportionment can be achieved by combining Factor Analysis with other statistical techniques. Factor Analysis/Multiple Regression (FA/MR) was developed in previous studies (Kleinman, 1977; Daisey *et al.*, 1981; Morandi , 1985) to quantitatively apportion contributions of pollution sources. In FA/MR, source patterns and specific source tracers are first identified by FA, and total PM mass is regressed against a unique tracer of each PM source using stepwise multiple regression. The mathematical expression of FA/MR is

$$Y = \sum K_i X_i + R,$$

where Y is the IPM concentration, K_i is the ith regression coefficient, X_i is atmospheric concentration of the source tracer for the ith source type, and R is a constant.

[FA/MR]/PR was used to estimate relative risks (or rate ratio, RR) of PM sources indirectly: individual PM sources were first identified by FA/MR, and then the individual source-specific IPM masses for all significant sources were then regressed against the mortality by PR to assess the quantitative relationship between source-specific IPM mass and mortality. For example, FA/PR identified three significant sources (oil burning, Zn/Cd processing and sulfate) associated with TDD in Newark. The RR estimation for each significant source is estimated by [FA/MR]/PR as follows:

1) by FA/MR: $IPM_{total\ mass} = IPM_{oil\ burning} + IPM_{ind\text{-}Zn} + IPM_{sulfate} + IPM_{dust} + IPM_{motor} + IPM_{ind\text{-}Cu} + IPM_{ind\text{-}Fe}$

2) by PR: $\log(\text{mortality}) = \beta_0 + \beta_1 * IPM_{oil\ burning} + \beta_2 * IPM_{ind\text{-}Zn} + \beta_3 * IPM_{sulfate} + \beta_4 * T_{avg}$

$RR_i = e^{(\beta i\ *\ change\ of\ source\text{-}specific\ IPM\ concentration)}$

(β_i: regression coefficient from PR).

RESULTS

Simple Poisson Model

In Newark, significant associations were found between three single exposure metrics - IPM, FPM, and sulfate - and TDD and CRDD ($p \leq 0.01$, Table 2). In Camden, IPM and FPM were significantly associated with TDD and all three single PM exposure metrics were all significantly associated with CRDD ($p \leq 0.05$). In Elizabeth, none of the three PM metrics were significantly associated with TDD or CRDD.

Of the three organic fractions, only cyclohexane-solubles (CX) was a significant predictor for TDD and CRDD in Newark and Camden. This organic fraction contains aliphatic hydrocarbons and polycyclic aromatic hydrocarbons, as well as other non-polar compounds. Acetone-solubles (ACE) was a significant predictor for TDD and CRDD only in Camden. No significant association was found between mortality and three organic fractions in Elizabeth.

The results of lag effect models, the increased mortality as a result of previous exposure, were comparable to the results of using concurrent day concentration in the simple Poisson models (details in Appendix A). Correlation analysis showed that the concentrations of individual pollutants on five consecutive days were highly correlated, as

might be expected ($p \leq 0.10$) and thus no difference could be found in the results of various lagged predictors (the concurrent day, 1-day lag and 2-day lag concentrations; 2 to 4 day averages) using simple Poisson models.

FA/PR

The results of Factor Analysis in this study were similar to Morandi's (1985), though the factor loadings were slightly different because we used an expanded dataset. The output of factor analysis on Newark data is shown in Table 3 as an example (detailed output of FA in three cities shown in Appendix B). Tracers used to identify PM sources are as follows (USEPA, 1990): V and Ni for oil burning sources; CO and Pb for motor emissions; Mn and Fe for dust; SO_4 for secondary aerosol and Zn, Cd, and Cu for various industrial sources. Statistically significant results of FA/PR using GEE, presented in Table 4, are discussed below by city (detailed results of FA/PR shown in Appendix C).

In Newark, oil burning sources (tracers: V and Ni), industrial sources (tracer: Zn and Cd) and sulfate aerosol showed positive associations with TDD. For CRDD, only sulfate was a significant source.

In Camden, oil-burning and motor vehicle emissions (tracers: Pb and CO) were two important sources for TDD. Sources traced by copper showed a negative association with TDD. Three PM sources were significant predictors for CRDD: oil burning, motor vehicles and sulfate.

In Elizabeth, resuspended dust (tracers: Fe and Mn) showed a negative association with TDD. Industrial sources traced by Cd showed a positive association with CRDD. Two other sources, resuspended dust and industrial sources traced by copper, showed negative associations with CRDD.

[FA/MR]/PR

Relative risks (RR) associated with each significant PM source were estimated indirectly by [FA/MR]/PR. This method apportioned source-specific PM masses by FA/MR first and then estimated the RRs of each IPM source. RRs obtained from [FA/MR]/PR and the conventional approach using total PM mass are presented in Table 5. These results suggest that FA/PR is a more sensitive approach for modeling mortality and provides more information about the possible causal relationships between PM chemical species exposure/PM sources and mortality than does the conventional approach using total IPM mass.

Model Comparison

We compared the traditional approach using a single PM exposure metric to the three variations of FA/PR (listed below) by using two statistical tests (described in Appendix D), the Akaike Information Criterion (AIC, Sakamoto *et al.*, 1986) and the Coefficient of Determination (COD, Nagelkerke, 1991).

Model 1 Traditional: $\log(\text{mortality}) = \beta_0 + \beta_1 * T_{avg} + \beta_2 * IPM_{mass}$

Model 2 FA/PR (a): $\log(\text{mortality}) = \beta_0 + \beta_1 * T_{avg} + \Sigma(\beta_{FSp} * \text{Factor Scores}_p)$,

Model 3 FA/PR (b): $\log(\text{mortality}) = \beta_0 + \beta_1 * T_{avg} + \Sigma(\beta_{FSq} * \text{Factor Scores}_q)$, a refined model from Model 2 in which only statistically significant factors were used.

Model 4 FA/PR (c): $\log(\text{mortality}) = \beta_0 + \beta_1 * T_{avg} + \Sigma(\beta_{Mq} * \text{single marker}_q)$, modified from Model 3.

where p is number of all PM sources resolved from FA, q is number of significant sources identified in the FA/PR, βs are the regression coefficients (β_{FSp} and β_{FSq} are regression coefficients for factor scores; β_{Mq} are coefficients for single marker). Single source tracers used for Model 4 were as follows: V for oil burning sources; SO_4 for sulfate aerosol; Pb for motor vehicle emission; Mn for geological sources.

The results of AICs, reported in detail in Appendix D, indicate that FA/PR (Model 2 or Model 3) is generally better than or equivalent to the conventional approach using total PM mass as an exposure metric (Model 1). In general, Model 3 is better than or equivalent to Model 2. This suggests that not all PM sources improve the fit of the model significantly; thus Model 2 with its larger number of PM source exposure metrics is no better than Model 3 which uses only PM sources identified by FA/PR as significantly

associated with mortality. The single marker model (Model 4), is no better than Model 3, depending on the number of significant factors and the source characteristics of those factors. Models 3 and Models 4 are comparable if the number of significant factors is small and those significant sources can be well represented by a unique tracer, such as sulfate. Based on the results of COD, Model 3 is better than Model 4 in all cases. This suggests that the use of factor scores is better than the use of simplified single marker for mortality prediction. Furthermore, using simplified exposure metrics, such as total PM mass or sulfate, may not be sufficient to capture the real relationship between PM exposure and adverse heath effects. Overall, Model 3 is better than or equivalent to the other three models.

DISCUSSION

Consistent with previous PM studies (Pope *et al.*, 1995), we found significant associations between single exposure metrics (IPM, FPM, and sulfate) and daily mortality in Newark and Camden although the ATEOS dataset is very small. The cyclohexane-soluble (CX) organic fraction was also a statistically significant predictor of mortality for Newark and Camden. ACE was significantly associated with mortality found only in Camden. Extractable organic matter (sum of CX, ACE, and DCM) accounts for about 30% of IPM mass in these three ATEOS sites, comparable to sulfate mass. The CX fraction accounts for about 9% and 7% of IPM mass in Newark and Camden, respectively. Analyses by Morandi (1985) indicate that the major sources of the non-polar CX-solubles in Newark are motor vehicles (26%), soil resuspension (21%), oil-burning (20%), and unidentified (34%).

With the FA/PR method, certain PM sources were found to be significant predictors for mortality: oil burning, sulfate aerosol, industry (traced by Zn/Cd), and motor vehicles. Information on local PM emission sources in each city ("qualitative microinventory", Morandi, 1985) and emissions profiles (of such PM sources) with known tracers (USEPA, 1990) were used to interpret the results of Factor Analysis, especially the industrial sources. Sources traced by zinc and cadmium were the Zn/Cd processing (e.g., smelters) located northeast of the sampling site in Newark; sources traced by copper were multiple industrial processing (e.g., fabricator, platers and illegal wire recovery operations) in Newark; sources traced by copper in Camden are incineration emissions. Oil burning sources, including industrial uses and residential heating, were scattered in space. PM from geological sources (tracers: Mn and Fe) or other industrial sources (tracers: Cu and Pb) were not significantly associated with increased mortality based on our analyses. Relative risks obtained from the FA/PR (or [FA/MR]/PR) are larger than those from the conventional approach using total PM mass (Table 5). This suggests that the chemical differences may play a role in the induced PM health effects and risks posed are not the same for various PM sources in the epidemiological settings.

The observed positive relationships between ambient PM sources and adverse health effects are supported by toxicological studies. For example, Holian *et al.* (1998) showed that urban particles and residual fly ash induced apoptosis of human alveolar macrophages, while Mt. St. Helen's volcanic ash showed no effects at all. Taking residual oil fly ash (ROFA) as an example, Costa *et al.* (1997) and Dreher *et al.* (1997) demonstrated that the lung dose of bioavailable transition metals (e.g., V, Ni and Zn) in

ROFA, not the total PM mass, is the primary determinant of the acute inflammatory response. Kodavanti *et al.* (1998) also found that *in vivo* ROFA-induced acute pulmonary inflammation was associated with its water-leachable V content and protein leakage was associated with its water-leachable Ni content. These results suggest that the potency and the mechanism of pulmonary injury will differ between emissions containing V and Ni.

In addition to the difference in source toxicity, those significant factors may also represent unmeasured chemical species or other parameters that are related to those PM sources and mortality. Those resolved significant sources possibly may serve as indicators of non-measured chemical species that are biologically important. Moreover, FA/PR cannot differentiate the effects of proximity to residents and toxicity of the PM sources. Those identified as significant sources of mortality may be in closer proximity to residents and thus have higher exposure effectiveness (EE, the amount of PM reaching human's respiratory tract divided by the amount of emissions from each PM source category). It is likely that EE may be different for the same PM source in three cities and thus different results of the same sources were found in FA/PR among the 3 cities. We have limited understanding in the real causal agent (or "the best exposure metric") for the increased mortality, but not all PM sources should be weighed equally for the adverse health effects found in this study.

Sulfate was used both as a single chemical species in the simple Poisson model and as a single PM source in FA/PR analyses. In Newark, sulfate was found to be a significant predictor for TDD in both types of analyses. Relative risks of sulfate obtained from both the simple Poisson model (the first analyses) and [FA/MR]/PR (the third analysis) were comparable and thus the use of FA/PR can be an alternative method to the

conventional analysis using total PM mass. Model comparisons using AIC and COD suggest that FA/PR may be a better tool for epidemiological studies if chemical speciation information is available.

The associations between mortality and ambient pollution in Elizabeth did not show patterns similar to those in Newark and Camden, even though Elizabeth is in close proximity to Newark and similar environmental concentrations were observed. A review of the population characteristics of the three study cities including socioeconomic status (SES, listed in Table 6) may help explain this difference. Elizabeth has the highest education levels (percentage of residents with 12 years of school completed or more: Elizabeth- 69 %; Newark- 58%; Camden- 66%) and income levels (percentage of residents below poverty level: Elizabeth- 8%; Newark- 18%; Camden- 12%) of the three cities. Studies (e.g., Lantz *et al.*, 1998) have shown that people with lower SES tend to have higher prevalence of behaviors that correlate with high health risk (e.g., cigarette smoking, alcohol drinking, and sedentary lifestyle). People with lower SES may be more likely to be in poor health and thus more susceptible to environmental stress. We hypothesize that the higher SES of the Elizabeth population may be the reason for the different pollution-mortality results in Elizabeth but appropriate data were not available to test the hypothesis.

Lag effects were evaluated in FA/PR models using an approach similar to that used in simple Poisson models. We found no consistent patterns in the lag structure of factor scores in FA/PR (details in Appendix C), in contrast to what we found for the simple Poisson models. This may be due to an implicit assumption for lag models that

health effects caused by all of the identified sources follow the same pattern, either decreasing or increasing, over several lagged days.

The detailed PM speciation measurements from the ATEOS study provided us an opportunity to investigate PM health effects by FA/PR. However, the small sample size limited the statistical power to detect the small change in daily mortality and made any further statistical methods to smooth the time series data impossible.

PM exposure metrics used in this study were obtained from ambient monitoring stations, as is conventional in epidemiological studies. Total human exposure may not be the same as the ambient concentrations. It was found that 75% of indoor PM_{10} and 83% of indoor $PM_{2.5}$ in Riverside, California homes was contributed by outdoor air (Ozkaynak *et al.*, 1996). In addition, indoor sources of PM, including environmental tobacco smoke, may also contribute to total PM exposure and the adverse health effects. However, tracers of the significant sources identified in this study (V, Ni, Zn, Cd) are mainly distributed in the small particles ($d_{50} < 2\mu m$, Lee *et al.*, 1973, Davidson *et al.*, 1986) that are more likely to penetrate indoors and thus have similar concentrations between indoor and outdoor environments, provided no presence of indoor sources.

The FA/PR method identifies important sources associated with mortality. However, because of the qualitative nature of factor analysis, quantification of risk associated with identified sources was estimated by [FA/MR]/PR. Since FA/MR uses a single tracer to apportion PM mass from each source, uncertainties are increased especially when tracers used are not unique among sources. Future development of risk quantification techniques will further enhance the usefulness of FA/PR in our understanding of PM.

In this exploratory study, FA/PR provides a means to extract more useful information about relationships between daily mortality and source-related exposure metrics composed of various chemical species than does the conventional use of total PM mass as an exposure metric. Using the FA/PR method, we have obtained the first epidemiological evidence indicating that chemical differences play a role in PM-induced health effects and that risks posed are not the same for PM sources with different characteristics. Future studies are recommended to apply the FA/PR to other larger-scale chemical speciation data to replicate and confirm the findings of this study.

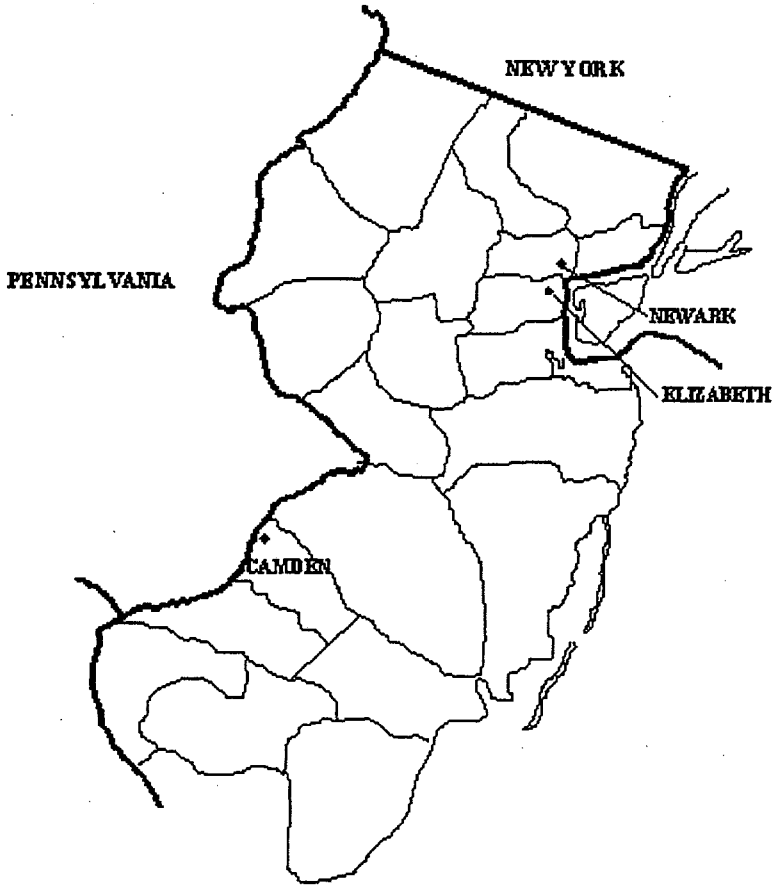

Figure 1. ATEOS SAMPLING SITES

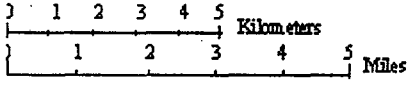

Figure 2. Newark data

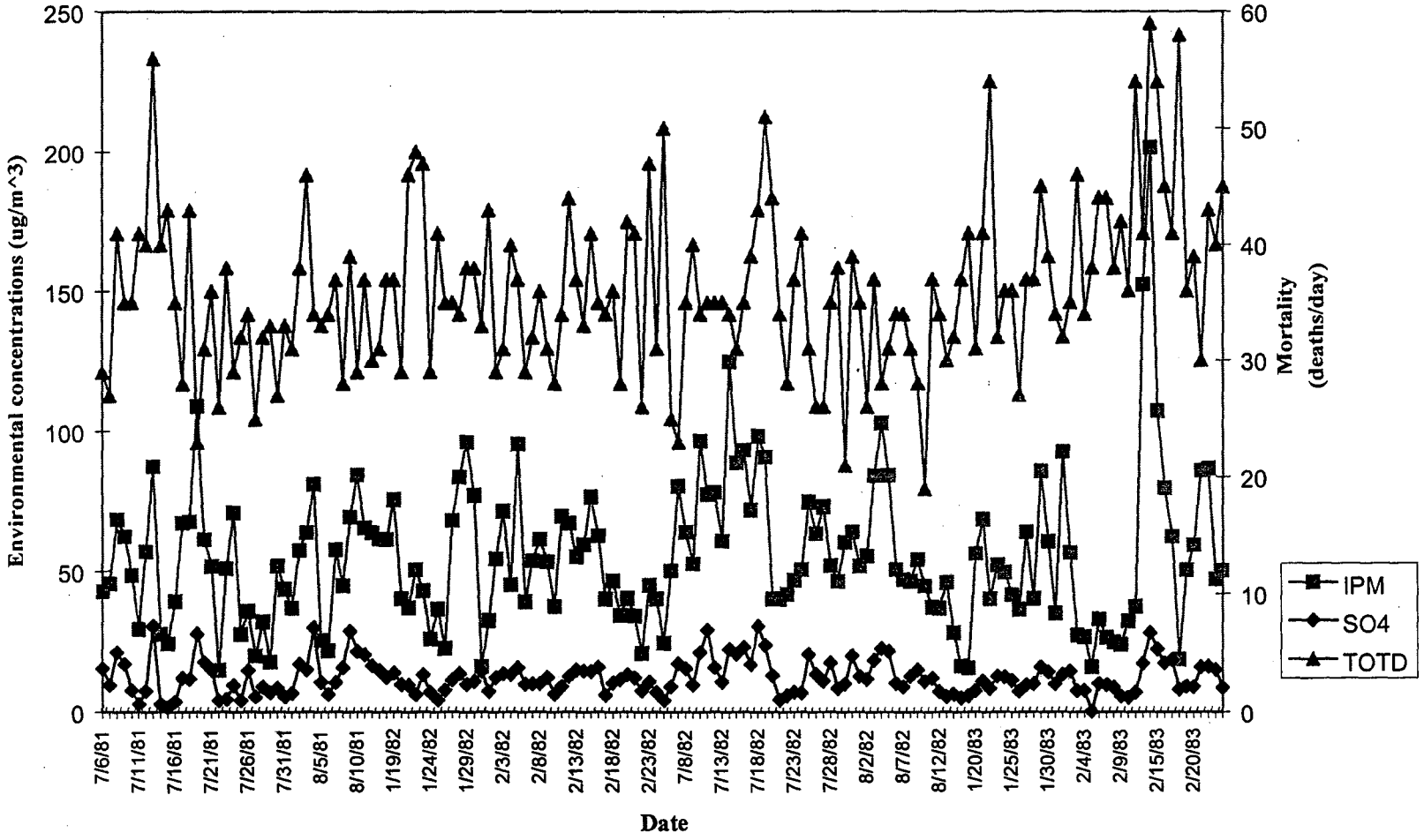

Figure 3. Camden data

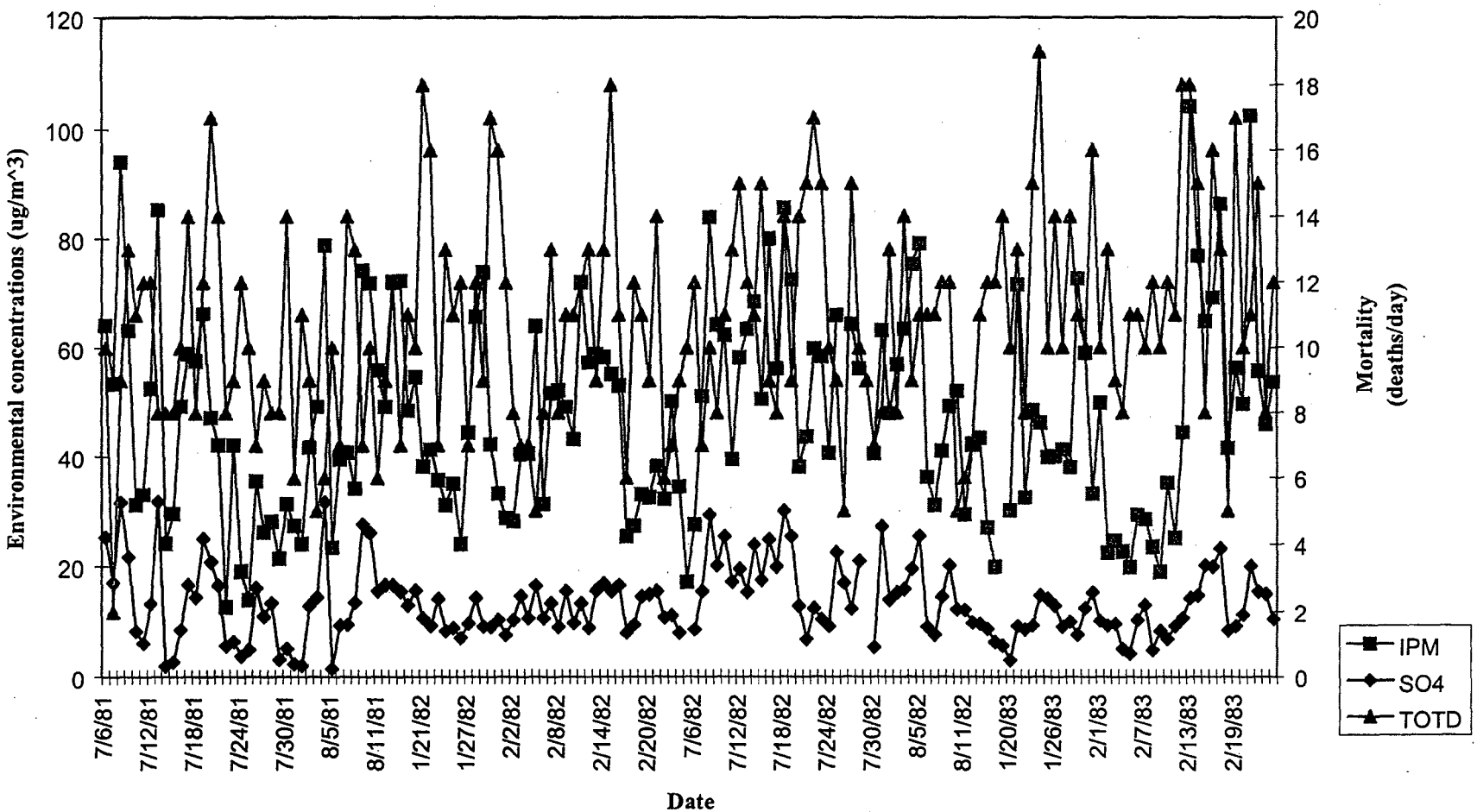

Figure 4. Elizabeth data

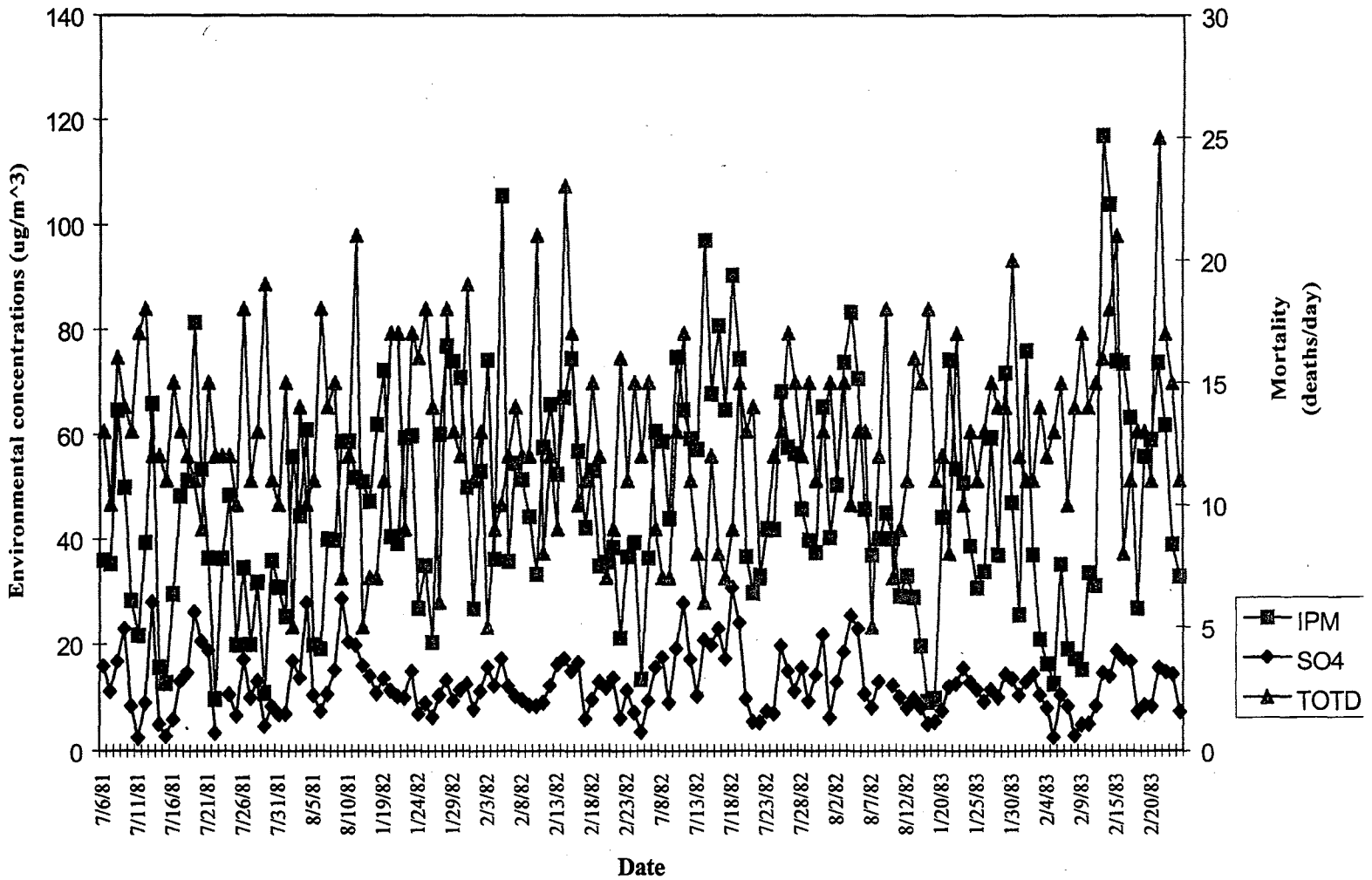

Figure 5. The FA/PR method.

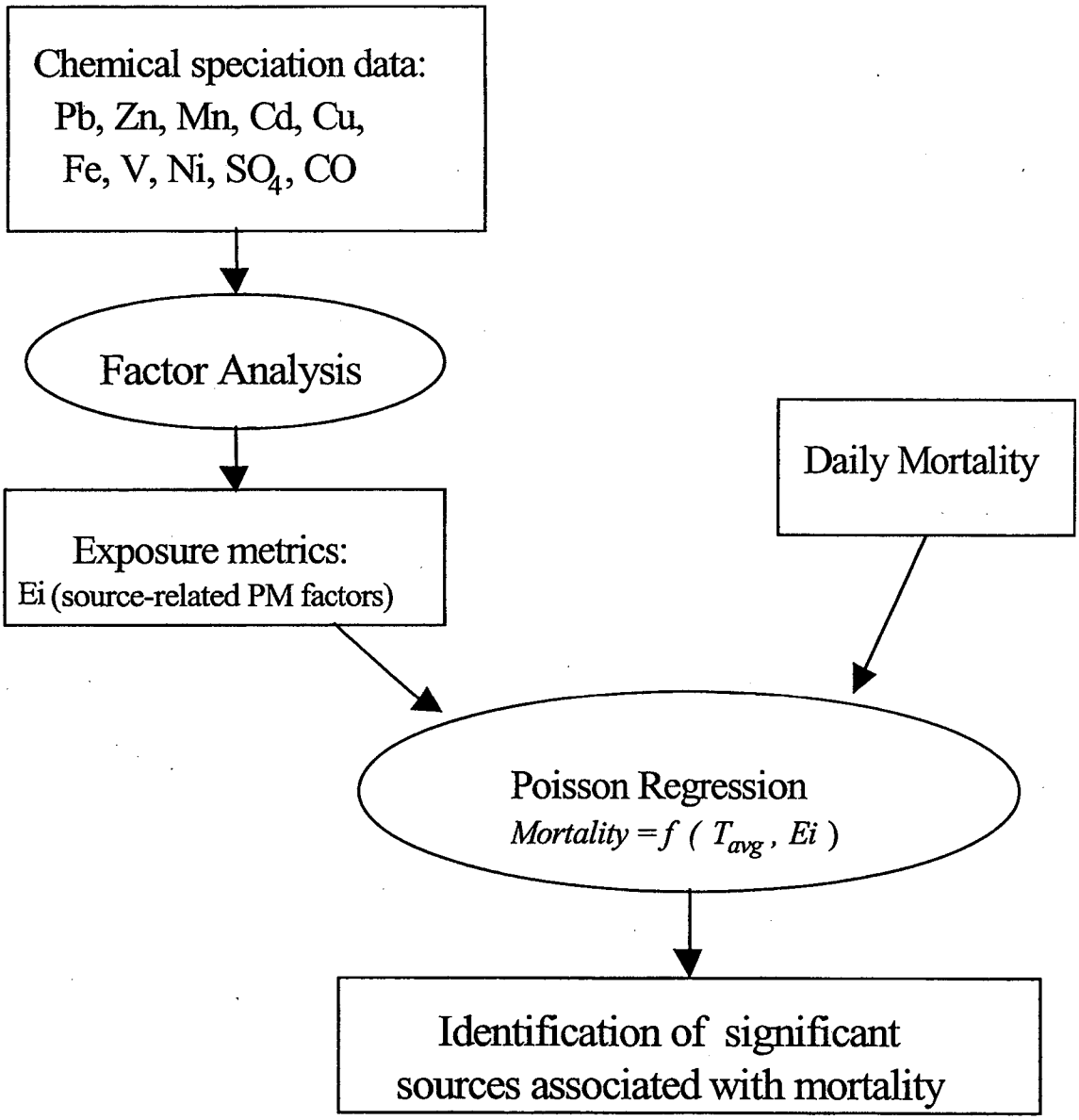

Table 1. Data summary for ATEOS sampling periods: arithmetic mean (standard deviation).

	Newark[1]	Elizabeth	Camden
IPM ($\mu g/m^3$)	55.5 (26.4)	47.0 (20.9)	47.5 (18.8)
FPM ($\mu g/m^3$)	42.1 (22.0)	37.1 (19.8)	39.9 (18.0)
Sulfate[2] ($\mu g/m^3$)	12.4 (6.2)	12.4 (5.8)	13.4 (6.5)
Pb (ng/m^3)	467 (335.6)	445.9 (337.5)	321.8 (203.6)
Mn (ng/m^3)	14.4 (10.3)	8.4 (7.3)	13.0 (12.3
Cu (ng/m^3)	41.2 (52.1)	48.9 (74.7)	22.1 (23.3)
V (ng/m^3)	44.2 (52.9)	42.3 (47.3)	35.8 (31.9)
Cd (ng/m^3)	7.2 (13.5)	3.1 (4.0)	1.9 (2.3)
Zn (ng/m^3)	876.8 (2744.8)	229.2 (413.8)	183.9 (155.4)
Fe (ng/m^3)	860.9 (424.7)	627.5 (377.1)	587.5 (352.3)
Ni (ng/m^3)	20.1 (16.3)	18.0 (16.8)	21.9 (16.0)
CX ($\mu g/m^3$)	4.9 (3.5)	4.8 (4.0)	3.4 (2.3)
DCM ($\mu g/m^3$)	1.4 (0.9)	1.2 (0.8)	1.2 (0.8)
ACE ($\mu g/m^3$)	9.6 (6.1)	7.8 (5.2)	8.7 (4.8)
CO_{max} (ppt)	3.5 (2.2)	3.2 (1.7)	2.6 (1.6)
T_{avg} (°F)	56.0 (23.8)	55.8 (24.2)	54.9 (23.9)
Population	1,408,088	504,094	471,650
TDD (deaths/day)	37 (8)	13 (4)	11 (3)
CRDD (deaths/day)	21 (6)	7 (3)	6 (3)

[1] The population characteristics (population number and mortality) in Newark are the summation of Essex and Hudson counties.
[2] Chemical species included in the FA/PR are underlined.

Table 2. Results of simple Poisson models: parameter estimate (standard error).

Site	TDD				CRDD			
	IPM	FPM	Sulfate	CX	IPM	FPM	Sulfate	CX
Newark	0.0011** (0.0001)	0.0017** (0.0003)	0.0061** (0.0018)	0.0057** (0.0017)	0.0015** (0.0004)	0.0020** (0.0004)	0.0085** (0.0009)	0.0026 (0.0014)
Camden	0.0021* (0.0010)	0.0022* (0.0011)	NS	0.0236** (0.0034)	0.0028** (0.0010)	0.0024* (0.0011)	0.0079** (0.0023)	0.0231** (0.0059)
Elizabeth	NS	NS	NS	NS	NS	NS	NS	NS

*: $p \leq 0.05$.

**: $p \leq 0.01$.

NS: not significant at $p \leq 0.10$.

Table 3. Results of factor analysis in Newark (source tracers underlined).

	Oil burning	Industrial-1	Geological	Industrial-2	Motor veh.	Sulfate	Industrial-3
	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	FACTOR7
Pb	0.37	0.78	0.21	0.08	0.35	0.00	0.01
Mn	0.13	0.13	0.95	0.14	0.10	0.12	-0.01
Cd	0.08	0.26	0.17	0.60	0.23	0.07	0.65
Cu	0.15	0.84	0.20	0.29	0.00	0.13	0.23
V	0.97	0.12	0.00	0.04	0.07	0.01	0.02
Zn	0.21	0.19	0.02	0.88	0.27	0.09	0.06
Fe	-0.01	0.33	0.71	-0.11	-0.01	0.20	0.50
Ni	0.82	0.28	0.18	0.28	0.20	0.04	0.05
SO_4	0.02	0.08	0.18	0.09	0.08	0.97	0.06
CO	0.21	0.17	0.09	0.34	0.87	0.10	0.10
Variance explained by each factor (%)							
	47.6	16.9	11.8	8.7	6.5	5.2	3.2

Table 4. Results of FA/PR analyses: p values (only p values ≤ 0.10 were reported).

ATEOS site	Newark		Camden		Elizabeth	
Source (tracer)	TDD	CRDD	TDD	CRDD	TDD	CRDD
Oil burning (V, Ni)	0.019[1]	NS	0.001	<0.001	NS	NS
Industrial -1[2] (Cu)	NS[3]	NS	0.056	NS	NS	0.030
Geological (Mn, Fe)	NS	NS	NS	NS	0.081	0.062
Industrial -2[4] (Zn)	<0.001	NS	NS	NS	NS	NS
Motor veh.[5](CO / CO, Pb)	NS	NS	<0.001	<0.001	NS	NS
Sulfate aerosol (SO_4)	0.002	<0.001	NS	0.055	NS	NS
Industrial -3 (Cd)	0.089	NS	NS	NS	NS	0.029

[1] P values less than 0.05 were boxed; negative and significant associations were underlined.
[2] Tracers for Ind-1 are Pb and Cu in Newark and Elizabeth; Cu only in Camden.
[3] NS: non-significant at p ≤ 0.10.
[4] Tracers for Ind-2 are Zn and Cd in Newark; Zn only in Elizabeth and Camden.
[5] Tracers for motor vehicles are CO and Pb in Camden; CO only in Newark and Elizabeth. Lead was used to be a unique marker for automobile emissions. With its phase-out in early 90s, lead from other non-auto sources became relatively more important during the transition period in Newark and Elizabeth.

Table 5. Relative risks associated with an increase of 10 $\mu g/m^3$ by the traditional approach using total PM mass and [FA/MR]/PR.

	Traditional approach (the first method)	[FA/MR]/PR (the third method)		
Newark	$IPM_{total\ mass}$	IPM_{oil}	$IPM_{ind\text{-}Zn}$	$IPM_{sulfate}$
TDD (total daily death)	1.01**	NS	1.03**	1.02**
CRDD (cardio-respiratory daily death)	1.02**	NA	NA	1.04**
Camden	$IPM_{total\ mass}$	IPM_{oil}	IPM_{motor}	$IPM_{sulfate}$
TDD (total daily death)	1.02*	1.11**	1.10*	NS
CRDD (cardio-respiratory daily death)	1.03**	1.12*	NS	1.02

*: $p < 0.05$ in [FA/MR]/PR.

**: $p < 0.01$ in [FA/MR]/PR.

NS: non-significant in [FA/MR]/PR at $p = 0.10$ level, but that source was significant in FA/PR.

NA: Non-significant sources in FA/PR at $p = 0.10$ level.

Table 6. Socioeconomic status and population data in New Jersey

(Source: 1980 Census data)

ATEOS counties	Union (Elizabeth)	Essex (Newark)	Hudson (Newark)	Camden
Total population (1980)	504,094	851,116	556,972	471,650
Population per square mile	4,885.6	6,695.9	11,993.4	2,112.5
Median age	34.7	31.5	32.3	30.5
Death rate (per 1,000 population)	9.9	10.1	11	9.4
% living in urban	100	100	100	95.1
Race				
White (%)	80.9	57.6	77.4	81.4
Black (%)	16.1	37.2	12.5	14.3
Others (%)	1.3	1.5	3.0	1.3
Years of school completed (persons 25 or older)				
% with 12 yr. or more	68.6	62.8	51.5	65.5
2% with 16 yr. or more	19.1	18	11.2	16.2
Median years of school completed	12.5	12.4	12	12.4
Income characteristics (1980)				
Personal income (per capita)	13,368	11,664	11,034	10,499
Median household income	25266	19931	17659	20998
% families below poverty level	5.8	15.2	14.7	9.6
% persons below poverty level	7.5	17.9	16.9	11.8
Children under 18 below poverty level	11.4	28.3	26.9	18.2

REFERENCES

Costa, D. L., and Dreher, K. L. 1997. Bioavailable transition metals in particulate matter mediate cardiopulmonary injury in healthy and compromised animal models. *Environ. Hea. Persp.* 105:1053-60.

Cox, W. M., and Clark, J. 1981. Ambient ozone concentration patterns among eastern U.S. urban Areas Using Factor Analysis. *J. Air Pollution Control. Asso.* 31:762-766.

Daisey, J. M., and Kneip, T. J. 1981. Atmospheric particulate organic matter-multivariate models for identifying sources and estimating their contributions to the ambient aerosol. In: *Atmospheric Aerosol: Source/ Air Quality Relationships.* Macias E.S., and Hopke P.K., eds. Symposium Series No. 167 pp. 197-221. American Chemistry Society, Washington D.C.

Daisey, J. M. 1987. Chemical composition of inhalable particulate matter-Seasonal and intersite comparison. *Toxic Air Pollution: A Comprehensive Study of Non-criteria Air Pollutants*. Lewis Publishers, Inc. (Eds. Lioy, P. J. and Daisey, J. M.) p. 47-65.

Davidson, C. I., and Osborn, J. F 1986. The sizes of airborne trace metal containing particles. In: *Toxic Metals in the Atmosphere.* Nriagu, J. O., and Davidson, C. I., Eds. John Wiley & Sons. New York.

Dreher, K. L., Jaskot, R. H., Lehmann, J. R., Richards, J. H., McGee, J. K., Ghio, A. J. and Costa, D. L. 1997. Soluble transition metals mediate residual oil fly ash induced acute lung injury. *J. Tox. Environ. Hea.* 50:285-305.

Holian, A., Hamilton, R. F., Morandi, M. T., Brown, S. D., and Li, L. 1998. Urban Particle-induced Apoptosis and Phenotypes Shifts in Human Alveolar Macrophages. *Environ. Hea. Persp.* 106(3): 127-132.

Hopke P. K. 1985. *Receptor Modeling in Environmental Chemistry.* Chapter 7. Factor Analysis. pp.155-197. John Wiley & Sons, Inc. New York.

Janssen, N. A. H., Hoek, G., Brunekreef, B., Harssema, H., Mensink, I., Zuidhof, A. 1998. Personal sampling of particles in adults: Relation among personal, indoor, and outdoor air concentrations. American Journal of Epidemiology, 147: 537-547.

Kleinman, M. T. 1977. *The apportionment of Sources of Airborne Particulate Matter.* Ph.D. Dissertation. New York University, N.Y.

Kodavanti, U. P., Hauser, R., Christiani, D. C., Meng, Z.H., McGee, J., Ledbetter, A., Richards, J., and Costa, D.L. 1998. Pulmonary responses to oil fly ash particles in the rat differ by virtue of their specific soluble metals. *Tox. Sci.* 43: 204-12.

Lantz, P. M., House, J. S., Lepkowski, J. M., Williams, D. R., Mero, R. P., and Chen, J. 1998. Socioeconomic factors, health behaviors, and mortality: results from a nationally representative prospective study of US adults. *J. Am. Med. Assoc.* 279: 1703-8.

Lee, R. E., and Lehmden, D. J. 1973. Trace metal pollution in the environment. J. Air Poll. Cont. Assoc. 23: 853-857.

Liang, K. Y., and Zeger, S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometricka* 73:13-22.

Lioy, P. J., Daisey, J. M., Greenberg, A., and Harkov, R. 1985. A Major Wintertime (1983) Pollution Episode in Northern New Jersey: Analyses of the Accumulation and Spatial Distribution of Inhalable Particulate Matter, Extractable Organic Matter and Other Species. *Atmos. Environ.* 19: 429-436.

Lioy, P. J. and Daisey, J. M. (editors). 1987. *Toxic Air Pollution: A Comprehensive Study of Non-criteria Air Pollutants*. Lewis Publishers, Inc. Chelsea, Michigan.

Morandi, M. T. 1985. *Development of Source Apportionment Models for Inhalable Particulate Matter and its Extractable Organic Fractions in Urban Areas of New Jersey*. Doctoral Dissertation, New York University, N.Y.

Morandi, M. T., Lioy, P. J., and Daisey, J. M. 1991. Comparison of Two Multivariate Modeling Approaches for the Source Apportionment of Inhalable Particulate Matter in Newark, NJ. *Atmos. Environ.* 25: 927-937.

Nagelkerke, N. J. D. 1991. A note on a general definition of the Coefficient of Determination. *Biometrika* 78: 691-692.

Ozkaynak, H., Xue, J., Spengler, J., Wallace, L., Pellizzari, E., and Jenkins, P. 1996. Personal exposure to airborne particles and metals- Results from the Particle TEAM Study in Riverside, California. *J. Exp. Ana. Environ. Epi.* 6:57-78.

Pope, C. A. III, Dockery, D. W., and Schwartz, J. 1995. Review of epidemiological evidence of health effects of particulate air pollution. *Inhal. Tox.* 7:1-18.

Sakamoto, Y., Ishiguro, M., and Kitagawa G. 1986. *Akaike Information Criterion Statistics.* Tokyo: KTK Scientific Publishers; Dordrecht; Boston: D. Reidel; Hingham, Ma: Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers.

SAS. 1989. *SAS/STAT User Guide.* Version 6, Fourth Edition, Volume 1. SAS Institute Inc., Cary, NC, USA.

United States Environmental Protection Agency. 1990. *Air Emissions Species Manual, Volume II: Particulate Matter Species Profiles,* EPA-450/2-90-001. January 1990. Research Triangle Park, NC.

Appendix A. Results of basic Poisson models: Total / Cardiorespiratory deaths = f (Tavg, IPM / FPM/ SO4)

1. Single day concentration with GEE:

Note: Only parameter estimates of p values less than 0.10 are listed; otherwise recorded as NA /NS (non significant).

Newark						
	TDD			CRDD		
1. IPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0011	0.0001	<0.0001	0.0015	0.0004	0.0009
1-d lag	0.0015	0.0009	0.0874	NA	NA	N.S.
2-d lag	0.0012	0.0003	0.0001	0.0020	0.0004	<0.0001
3-d lag	NA	NA	N.S.	0.0070	0.0002	0.0003
2. FPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0017	0.0003	<0.0001	0.0020	0.0004	<0.0001
1-d lag	0.0023	0.0009	0.0074	0.0028	0.0011	0.0097
2-d lag	0.0016	0.0002	<0.0001	0.0028	0.0004	<0.0001
3-d lag	N.S.	NA	N.S.	0.0013	0.0005	0.0084
3. Sulfate	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0061	0.0018	0.0007	0.0085	0.0009	<0.0001
1-d lag	NA	NA	N.S.	NA	NA	N.S.
2-d lag	0.0013	0.0006	0.0212	0.0045	0.0017	0.0084
3-d lag	NA	NA	N.S.	NA	NA	N.S.

Elizabeth						
	TDD			CRDD		
1. IPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	NA	NA	N.S.	NA	NA	N.S.
1-d lag	NA	NA	N.S.	0.0032	0.0013	0.0125
2-d lag	NA	NA	N.S.	NA	NA	N.S.
3-d lag	NA	NA	N.S.	NA	NA	N.S.
2. FPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	NA	NA	N.S.	NA	NA	N.S.
1-d lag	NA	NA	N.S.	0.0024	0.0012	0.0490
2-d lag	0.0016	0.0007	0.0301	NA	NA	N.S.
3-d lag	NA	NA	N.S.	NA	NA	N.S.
3. Sulfate	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	NA	NA	N.S.	NA	NA	N.S.
1-d lag	NA	NA	N.S.	NA	NA	N.S.
2-d lag	NA	NA	N.S.	-0.0079	0.0025	0.0013
3-d lag	NA	NA	N.S.	NA	NA	N.S.

Camden						
	TDD			CRDD		
1. IPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0021	0.0010	0.0369	0.0028	0.0010	0.0037
1-d lag	0.0013	0.0008	0.0958	0.0033	0.0013	0.0095
2-d lag	N.S.	NA	N.S.	*0.0036*	*0.0022*	*0.1012*
3-d lag	0.0041	0.0012	0.0006	0.0040	0.0019	0.0295
2. FPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0022	0.0011	0.0488	0.0024	0.0011	0.0276
1-d lag	0.0017	0.0003	<0.0001	0.0033	0.0008	0.0001
2-d lag	N.S.	NA	N.S.	*0.0034*	*0.0021*	*0.1063*
3-d lag	0.0047	0.0012	0.0001	0.0036	0.0016	0.0299
3. Sulfate	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	N.S.	NA	N.S.	0.0079	0.0023	0.0007
1-d lag	N.S.	NA	N.S.	0.0095	0.0048	0.0450
2-d lag	N.S.	NA	N.S.	0.0098	0.0042	0.0187
3-d lag	0.0074	0.0035	0.0343	0.0100	0.0032	0.0019

2. Results of moving average: Total death = Tavg + IPM / FPM/ SO4

Note: N-day moving average equals to the arithmetic mean of concurrent day concentration up to the (N-1) lag day concentration.

Newark						
	TDD			CRDD		
1. IPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0011	0.0001	<0.0001	0.0015	0.0004	0.0009
2-d moving avg.	0.0018	0.0006	0.0033	0.0022	0.0006	0.0001
3-d moving avg.	0.0023	0.0006	0.0003	0.0032	0.0006	<0.0001
4-d moving avg.	0.0023	0.0006	0.0001	0.0041	0.0007	<0.0001
2. FPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0017	0.0003	<0.0001	0.0020	0.0004	<0.0001
2-d moving avg.	0.0026	0.0008	0.0019	0.0032	0.0008	0.0001
3-d moving avg.	0.0032	0.0006	<0.0001	0.0045	0.0008	<0.0001
4-d moving avg.	0.0030	0.0008	0.0001	0.0049	0.0013	0.0001
3. Sulfate	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0061	0.0018	0.0007	0.0085	0.0009	<0.0001
2-d moving avg.	NA	NA	N.S.	0.0098	0.0060	0.1039
3-d moving avg.	0.0079	0.0046	0.0864	0.0126	0.0067	0.0616
4-d moving avg.	NA	NA	N.S.	0.0156	0.0094	0.0969

Elizabeth						
	TDD			CRDD		
1. IPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	NA	NA	N.S.	NA	NA	N.S.
2-d moving avg.	NA	NA	N.S.	NA	NA	N.S.
3-d moving avg.	NA	NA	N.S.	NA	NA	N.S.
4-d moving avg.	NA	NA	N.S.	NA	NA	N.S.
2. FPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	NA	NA	N.S.	NA	NA	N.S.
2-d moving avg.	NA	NA	N.S.	NA	NA	N.S.
3-d moving avg.	NA	NA	N.S.	NA	NA	N.S.
4-d moving avg.	NA	NA	N.S.	NA	NA	N.S.
3. Sulfate	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	NA	NA	N.S.	NA	NA	N.S.
2-d moving avg.	NA	NA	N.S.	NA	NA	N.S.
3-d moving avg.	NA	NA	N.S.	NA	NA	N.S.
4-d moving avg.	-0.0076	0.0022	0.0006	NA	NA	N.S.

Camden						
	TDD			CRDD		
1. IPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0021	0.0010	0.0369	0.0028	0.0010	0.0037
2-d moving avg.	0.0028	0.0004	<0.0001	0.0044	0.0005	<0.0001
3-d moving avg.	0.0036	0.0011	0.0007	0.0057	0.0020	0.0038
4-d moving avg.	0.0063	0.0019	0.0012	0.0078	0.0032	0.0149
2. FPM	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	0.0022	0.0011	0.0488	0.0024	0.0011	0.0276
2-d moving avg.	0.0026	0.0008	0.0006	0.0040	0.0004	<0.0001
3-d moving avg.	0.0031	0.0005	<0.0001	0.0050	0.0013	0.0002
4-d moving avg.	0.0055	0.0008	<0.0001	0.0060	0.0023	0.0089
3. Sulfate	parameter estimate	standard error	p value	parameter estimate	standard error	p value
concurrent day	NA	NA	N.S.	0.0079	0.0023	0.0007
2-d moving avg.	NA	NA	N.S.	0.0131	0.0043	0.0026
3-d moving avg.	0.0026	0.0009	0.0032	0.0193	0.0065	0.0031
4-d moving avg.	0.0097	0.0019	<0.0001	0.0262	0.0090	0.0037

Appendix B. Output of Factor Analysis

1. Newark

7 factors, Newark

Rotation Method: Varimax

Rotated Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	FACTOR7	
PB	0.36777	0.77793	0.20524	0.08328	0.35095	0.00254	0.01348	PB
MN	0.12991	0.13322	0.94608	0.13565	0.10487	0.12444	-0.00961	MN
CD	0.08330	0.26155	0.16970	0.59784	0.23101	0.07373	0.64938	CD
CU	0.15190	0.84474	0.19528	0.28735	0.00481	0.13240	0.23255	CU
V	0.96860	0.11762	-0.00294	0.04389	0.07425	0.00538	0.01838	V
ZN	0.21428	0.19338	0.02368	0.87644	0.26975	0.09039	0.05907	ZN
FE	-0.01121	0.33323	0.71154	-0.11066	-0.00615	0.20092	0.50059	FE
NI	0.82193	0.28088	0.18348	0.28309	0.19812	0.04226	0.04567	NI
SO4	0.02311	0.08033	0.17808	0.08713	0.07908	0.96997	0.06056	SO4
CO	0.20744	0.17144	0.08768	0.33522	0.87495	0.10245	0.09853	CO

Variance explained by each factor

FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	FACTOR7
1.885505	1.681937	1.584049	1.447746	1.076906	1.040150	0.745919

Final Communality Estimates: Total = 9.462212

PB	MN	CD	CU	V	ZN	FE	NI	SO4	CO
0.912842	0.974668	0.942050	0.929000	0.959829	0.936452	0.920696	0.911398	0.997059	0.978218

2. Camden

7 factors, Camden

Rotation Method: Varimax

Rotated Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	FACTOR7	
PB	0.40275	0.34624	0.60843	0.35667	0.05899	-0.01007	0.15226	PB
MN	0.14870	0.88809	0.00950	-0.06391	0.20714	-0.01075	0.17645	MN
CD	0.13951	0.06343	0.00259	0.88908	0.26841	-0.05660	0.21187	CD
CU	0.13365	0.14452	0.09387	0.24865	0.92953	0.00744	0.06791	CU
V	0.92195	0.20869	0.11064	0.14316	0.09262	0.08116	0.04903	V
ZN	0.26757	0.10500	0.22404	0.24527	0.07710	-0.06013	0.87610	ZN
FE	0.23430	0.84844	0.20337	0.18758	-0.02437	0.11675	-0.05136	FE
NI	0.84235	0.19284	0.23192	0.04922	0.10585	0.01177	0.33601	NI
SO4	0.06348	0.06748	-0.01880	-0.04553	0.00497	0.99237	-0.04331	SO4
CO	0.13279	0.05835	0.93315	-0.06820	0.07433	-0.01733	0.14493	CO

Variance explained by each factor

FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	FACTOR7
1.929359	1.753062	1.407806	1.108582	1.014328	1.012554	1.012196

Final Communality Estimates: Total = 9.237888

PB	MN	CD	CU	V	ZN	FE	NI	SO4	CO
0.806267	0.889145	0.934103	0.978077	0.943854	0.970090	0.868148	0.927200	0.997716	0.923289

3. Elizabeth

7 factors, Elizabeth

Rotated Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	FACTOR7	
PB	0.39478	0.37012	0.64908	0.23513	0.01802	0.25424	0.23874	PB
MN	0.33702	0.81181	0.18982	0.16609	-0.00535	0.12566	0.03945	MN
CD	0.14566	0.19801	0.29362	0.13198	0.01622	0.88678	0.21880	CD
CU	0.14370	0.17542	0.89036	0.14219	0.02034	0.21440	0.23002	CU
V	0.92258	0.25171	0.11973	0.04012	-0.01140	0.10885	0.03404	V
ZN	0.20332	0.16710	0.33098	0.12583	-0.01356	0.23651	0.85975	ZN
FE	0.21588	0.85447	0.15604	-0.05308	0.18183	0.12130	0.16189	FE
NI	0.86639	0.25880	0.19096	0.10540	-0.05570	0.07239	0.22130	NI
SO4	-0.04973	0.11557	0.01558	0.14347	0.97782	0.01163	-0.00966	SO4
CO	0.09885	0.05854	0.17824	0.94715	0.15929	0.11100	0.09969	CO

Variance explained by each factor

FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	FACTOR7
2.013274	1.771168	1.553009	1.069546	1.019005	1.012971	0.984880

Final Communality Estimates: Total = 9.423854

PB	MN	CD	CU	V	ZN	FE	NI	SO4	CO
0.891400	0.853607	0.998576	0.963674	0.943587	0.989932	0.877880	0.922502	0.993009	0.989686

Appendix C. Results of FA/PR.

Note: only results with p values less than 0.10 are recorded; N.S.: non significant at p = 0.10 level.

1. Newark data

a. Total daily death (TDD)

		Tavg	Factor 1 (V/Ni)	Factor 2 (Cu/Pb)	Factor 3 (Mn/Fe)	Factor 4 (Zn/Cd)	Factor 5 (CO/Pb)	Factor 6 (SO4)	Factor 7 (Fe/Cd)
1. FA/PR w/o GEE	beta (stand. err.)	-0.0022 (0.0008)	0.0261 (0.0151)			0.0353 (0.0128)		0.0346 (0.0152)	
	p value	0.0079	0.0842			0.0059		0.0233	
2. FA/PR with GEE	beta (stand. err.)	-0.0023 (0.0008)	0.0225 (0.0096)			0.0328 (0.0033)		0.0350 (0.0093)	-0.0191(0.0113)
	p value	0.0028	0.0191			<0.0001		0.0002	0.0895
3. Only significant factors w/ GEE	beta (stand. err.)	-0.0025 (0.0010)	0.0204 (0.0089)			0.0302 (0.0052)		0.0359 (0.0109)	
	p value	0.0144	0.0214			<0.0001		0.0010	
4. 1-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0013 (0.0009)	0.0478 (0.0094)			0.0408 (0.0023)		0.0059 (0.0260)	
	p value	N.S./ 0.1602	<0.0001			<0.0001		N.S./ 0.8204	
5. 2-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0022 (0.0006)	0.0121 (0.0068)			0.0345 (0.0060)		0.0075 (0.0113)	
	p value	0.0002	0.0767			<0.0001		N.S./ 0.5032	
6. 3-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0023 (0.0010)	-0.0054 (0.0051)			0.0039 (0.0029)		-0.0069 (0.0251)	
	p value	0.0200	N.S./ 0.2925			N.S./ 0.1814		N.S./ 0.7828	

b. Cardio-respiratory daily death (CRDD)

		Tavg	Factor 1 (V/Ni)	Factor 2 (Cu/Pb)	Factor 3 (Mn/Fe)	Factor 4 (Zn/Cd)	Factor 5 (CO/Pb)	Factor 6 (SO4)	Factor 7 (Fe/Cd)
1. FA/PR w/o GEE	beta (stand. err.)	-0.0028 (0.0011)				0.0297 (0.0167)		0.0502 (0.0197)	
	p value	0.0072				0.0754		0.0107	
2. FA/PR with GEE	beta (stand. err.)	-0.0039 (0.0004)						0.0540 (0.0079)	
	p value	<0.0001						<0.0001	
3. Only significant factors w/ GEE	beta (stand. err.)	-0.0032 (0.0006)						0.0488 (0.0093)	
	p value	<0.0001						<0.0001	
4. 1-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0024 (0.0010)						-0.0019 (0.0357)	
	p value	0.0184						N.S./ 0.9582	
5. 2-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0025 (0.0005)						0.0327 (0.0075)	
	p value	<0.0001						<0.0001	
6. 3-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0024 (0.0007)						0.0084 (0.0247)	
	p value	0.0013						N.S./ 0.7323	

2. Camden data

a. Total daily death (TDD)

		Tavg	Factor 1 (V/Ni)	Factor 2 (Mn/Fe)	Factor 3 (CO/Pb)	Factor 4 (Cd)	Factor 5 (Cu)	Factor 6 (SO4)	Factor 7 (Zn)
1. FA/PR w/o GEE	beta (stand. err.)		0.0761 (0.0303)		0.0581 (0.0240)				
	p value		0.0119		0.0153				
2. FA/PR with GEE	beta (stand. err.)		0.0765 (0.0238)		0.0577 (0.0164)		-0.0243 (0.0127)		
	p value	N.S.	0.0013		0.0004		0.0561		
3. Only significant factors w/ GEE	beta (stand. err.)		0.0714 (0.0208)		0.0554 (0.0161)				
	p value		0.0006		0.0006				
4. 1-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0024 (0.0011)	-0.0215 (0.0168)		0.0226 (0.0063)				
	p value	0.0250	N.S./ 0.2005		0.0004				
5. 2-d lag on significant factors w/ GEE	beta (stand. err.)		0.0267 (0.0179)		-0.0059 (0.0119)				
	p value		N.S./ 0.1347		N.S./ 0.6192				
6. 3-d lag on significant factors w/ GEE	beta (stand. err.)		0.0729 (0.0129)		-0.0035 (0.0086)				
	p value		<0.0001		N.S./ 0.6813				

b. Cardio-respiratory daily death (CRDD)

		Tavg	Factor 1 (V/Ni)	Factor 2 (Mn/Fe)	Factor 3 (CO/Pb)	Factor 4 (Cd)	Factor 5 (Cu)	Factor 6 (SO4)	Factor 7 (Zn)
1. FA/PR w/o GEE	beta (stand. err.)		0.0680 (0.0391)						
	p value	N.S.	0.0819						
2. FA/PR with GEE	beta (stand. err.)	-0.0018 (0.0009)	0.0664 (0.0197)		0.0381 (0.0101)			0.0248 (0.0129)	
	p value	0.0537	0.0007		0.0002			0.0550	
3. Only significant factors w/ GEE	beta (stand. err.)	-0.0024 (0.0012)	0.0589 (0.0173)		0.0372 (0.0084)			0.0294 (0.0134)	
	p value	0.0430	0.0007		<0.0001			0.0287	
4. 1-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0050 (0.0013)	-0.0229 (0.0215)		0.0253 (0.0023)			0.0581 (0.0389)	
	p value	0.0001	N.S./ 0.2868		<0.0001			N.S./ 0.1354	
5. 2-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0036 (0.0020)	0.0226 (0.0307)		-0.0200 (0.0147)			0.0485 (0.0238)	
	p value	0.0699	N.S./ 0.4628		N.S./ 0.1738			0.0415	
6. 3-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0033 (0.0015)	0.0574 (0.0193)		-0.0249 (0.0168)			0.0443 (0.0309)	
	p value	0.0279	0.0030		N.S./ 0.1385			N.S./ 0.1515	

3. Elizabeth data

a. Total daily death (TDD)

		Tavg	Factor 1 (V/Ni)	Factor 2 (Fe/Mn)	Factor 3 (Cu/Pb)	Factor 4 (CO)	Factor 5 (SO4)	Factor 6 (Cd)	Factor 7 (Zn)
1. FA/PR w/o GEE	beta (stand. err.)								
	p value	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
2. FA/PR with GEE	beta (stand. err.)			-0.0296 (0.0170)					
	p value	N.S.		0.0809					
3. Only significant factors w/ GEE	beta (stand. err.)	-0.0026 (0.0010)		-0.0276 (0.0158)					
	p value	0.0127		0.0807					
4. 1-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0026 (0.0007)		-0.0133 (0.0287)					
	p value	0.0004		N.S./ 0.6417					
5. 2-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0023 (0.0011)		-0.0206 (0.0547)					
	p value	0.0323		N.S./ 0.7070					
6. 3-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0024 (0.0008)		-0.0078 (0.0128)					
	p value	0.0022		N.S./ 0.5397					

b. Cardio-respiratory daily death (CRDD)

		Tavg	Factor 1 (V/Ni)	Factor 2 (Fe/Mn)	Factor 3 (Cu/Pb)	Factor 4 (CO)	Factor 5 (SO4)	Factor 6 (Cd)	Factor 7 (Zn)
1. FA/PR w/o GEE	beta (stand. err.)								
	p value	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
2. FA/PR with GEE	beta (stand. err.)			-0.0335 (0.0180)	-0.0284 (0.0131)			0.0051 (0.0023)	
	p value	N.S.		0.0622	0.0300			0.0288	
3. Only significant factors w/ GEE	beta (stand. err.)	-0.0021 (0.0010)		-0.0258 (0.0151)	-0.0283 (0.0208)			0.0081 (0.0035)	
	p value	0.0315		0.0868	N.S./ 0.1728			0.0208	
4. 1-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0024 (0.0009)		0.0358 (0.0258)	-0.0026 (0.0370)			0.0218 (0.0158)	
	p value	0.0082		N.S./ 0.1651	N.S./ 0.9439			N.S./ 0.1692	
5. 2-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0021 (0.0015)		0.0114 (0.0619)	0.0077 (0.0408)			-0.0294 (0.0159)	
	p value	N.S./ 0.1545		N.S./ 0.8541	N.S./ 0.8497			0.0651	
6. 3-d lag on significant factors w/ GEE	beta (stand. err.)	-0.0018 (0.0006)		-0.0076 (0.0243)	-0.0375 (0.0195)			-0.0478 (0.0143)	
	p value	0.0023		N.S./ 0.7538	0.0548			0.0008	

Appendix D. Model Comparison

In addition to the traditional simple Poisson model using PM mass concentration as the exposure metric, we considered three variations of FA/PR models. Four models of interest for comparison are listed in Table E-1. It is not straightforward to compare these models since each model has a different number of dependent variables/predictors and missing values. Two approaches, Akaike Information Criterion (AIC) and Coefficient of Determination, were applied to compare the models.

The Akaike Information Criterion (AIC) was first proposed in 1973 as a criterion for model selection. The goodness of fit for a specific model is measured by the mean expected log likelihood; the larger, the better. The mean expected log likelihood is estimated by the maximum log likelihood (MLL). MLL has a general tendency to overestimate the true value of the mean expected log likelihood and this tendency is more prominent for models with a larger number of parameters. This implies that if we select the model with the largest maximum log likelihood, a model with an unnecessarily large number of parameters is likely to be chosen. The AIC adjusts for the effects of the number of parameters on the log likelihood.

AIC= - 2 x (maximum log likelihood of the model)

+ 2 x (number of parameters of the model)

The model with the minimum AIC is considered to be the "best" model (subjective, not based on statistical tests). The difference of AIC values matters, not the actual values themselves. If the difference between the AICs for Model (m) and Model (n) is larger than 1-2, then it is considered to be significant. If $|AIC(m)-AIC(n)| \leq 1$,

then the goodness of fits of these two models are almost the same (Sakamoto *et al.*, 1986).

To further compare Model 3 and Model 4, which have the same number of parameters, we used the Coefficient of Determination (COD, Nagelkerke, 1991), the pseudo R^2.

All missing values of the parameters used in the models of interest for comparison (Table D-1) were deleted before calculating AICs by SPLUS. The same data set is used for all models evaluated by AIC. The parameter estimates reported in the text, however, may be based on a larger number of observations than are in that data set, depending on the number of missing values in the predictors for each model. As shown in Table D-2, FA/PR (Model 2 or Model 3) is generally better or equivalent than the conventional approach using PM total mass as exposure metric (Model 1). The AICs of Model 2, which uses factor scores of all PM sources as mortality predictors, are usually higher than those of Model 3. This suggests that not all PM sources are related to increased daily mortality; thus Model 2 with its larger number of PM source predictors is no better than Model 3 which uses only PM sources identified by FA/PR as significantly associated with mortality. The single marker model (Model 4), which uses a unique tracer to represent each source, is equivalent to or less effective than Model 3, depending on the number of significant factors and the source characteristics of those factors. Generally speaking, Model 3 and Model 4 give similar AICs if the number of significant factors is small and if those significant sources can be well represented by a single marker/tracer, such as sulfate.

Model 3 is better than Model 4 in all cases according to COD analysis (Table D-3). This suggests that the use of factor scores, composite measures of all chemical species in the factors, is better than the use of single marker for mortality prediction even though the underlying biological mechanism of multiple chemical speciation exposure is not clear at this point. Furthermore, using simplified exposure metrics, such as total PM mass or sulfate, may not be sufficient to characterize/capture the real relationship between PM exposure and adverse heath effects.

TABLE D-1. Four models of interest for comparison.

Simple Poisson Model	Mortality $= f(T_{avg}, IPM_{total\ mass})$	(1)
FA/PR (a)	Mortality $= f(T_{avg}, (\text{Factor scores})_k)$, k= number of all sources resolved from FA	(2)
FA/PR (b)	Mortality $= f(T_{avg}, (\text{Factor scores})_l)$, l= number of significant sources identified in the FA/PR, a refined model from Model (2).	(3)
FA/PR (c)	Mortality $= f(T_{avg}, (\text{Single marker}^1)_l)$, l = number of significant sources found in FA/PR, a modified model from Model (3).	(4)

[1] Examples of single tracer used for each source: V for oil burning; SO_4 for sulfate aerosol; Mn for dust; Pb-motor for motor vehicle emissions.

Table D-2. Model comparison by AIC.

	Dependent variable	Model 1	Model 2	Model 3	Model 4
Newark	Total daily death (TDD)	191.7	187.7	183.4	188.4
	Cardio-respiratory daily death (CRDD)	184.3	189.0	182.9	183.4
Camden	Total daily death (TDD)	141.5	142.9	135.0	136.4
	Cardio-respiratory daily death (CRDD)	138.8	146.6	138.2	136.7

Table D-3. Model comparison by Coefficient of Determination.

Dependent variable	COD	Model 3	Model 4
Newark – TDD	R^2	0.228	0.185
	Max-rescaled R^2	0.299	0.246
Newark – CRDD	R^2	0.423	0.216
	Max-rescaled R^2	0.549	0.286
Camden – TDD	R^2	0.438	0.393
	Max-rescaled R^2	0.569	0.522
Camden – CRDD	R^2	0.433	0.403
	Max-rescaled R^2	0.563	0.535

Ernest Orlando Lawrence Berkeley National Laboratory
One Cyclotron Road | Berkeley, California 94720

Prepared for the U.S. Department of Energy under Contract No. DE-AC03-76SF00098