



RESEARCH REPORT

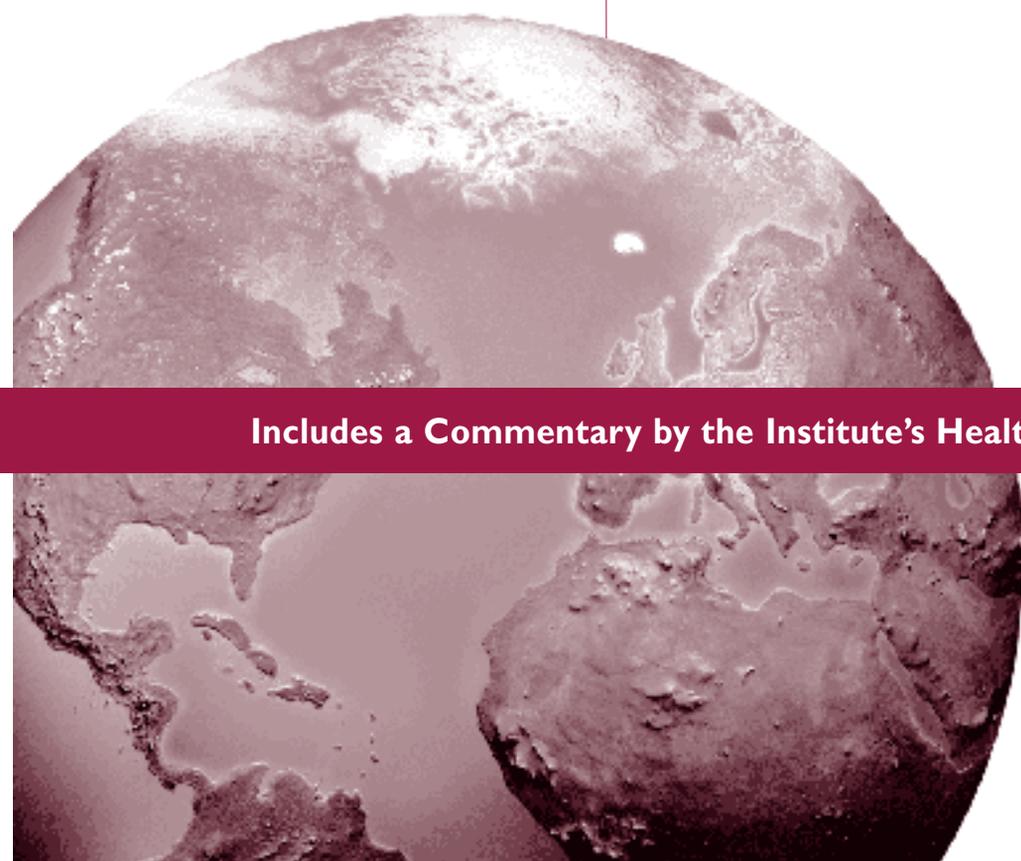
HEALTH  
EFFECTS  
INSTITUTE

Number 94 Part IV  
September 2005

# **The National Morbidity, Mortality, and Air Pollution Study**

## **Part IV: Hierarchical Bivariate Time-Series Models—A Combined Analysis of PM<sub>10</sub> Effects on Hospitalization and Mortality**

Francesca Dominici, Antonella Zanobetti,  
Scott L Zeger, Joel Schwartz, and Jonathan M Samet



Includes a Commentary by the Institute's Health Review Committee



## HEALTH EFFECTS INSTITUTE

The Health Effects Institute was chartered in 1980 as an independent and unbiased research organization to provide high quality, impartial, and relevant science on the health effects of emissions from motor vehicles, fuels, and other environmental sources. All results are provided to industry and government sponsors, other key decisionmakers, the scientific community, and the public. HEI funds research on all major pollutants, including air toxics, diesel exhaust, nitrogen oxides, ozone, and particulate matter. The Institute periodically engages in special review and evaluation of key questions in science that are highly relevant to the regulatory process. To date, HEI has supported more than 220 projects at institutions in North America, Europe, and Asia and has published over 160 Research Reports and Special Reports.

Typically, HEI receives half of its core funds from the US Environmental Protection Agency and half from 28 worldwide manufacturers and marketers of motor vehicles and engines who do business in the United States. Other public and private organizations periodically support special projects or certain research programs. Regardless of funding sources, HEI exercises complete autonomy in setting its research priorities and in reaching its conclusions.

An independent Board of Directors governs HEI. The Institute's Health Research Committee develops HEI's five-year Strategic Plan and initiates and oversees HEI-funded research. The Health Review Committee independently reviews all HEI research and provides a Commentary on the work's scientific quality and regulatory relevance. Both Committees draw distinguished scientists who are independent of sponsors and bring wide-ranging multidisciplinary expertise.

The results of each project and its Commentary are communicated widely through HEI's home page, Annual Conference, publications, and presentations to professional societies, legislative bodies, and public agencies.

# HEI STATEMENT

Synopsis of Research Report 94 Part IV

## The National Morbidity, Mortality, and Air Pollution Study, Part IV: Combined Analysis of PM<sub>10</sub> Effects on Hospitalization and Mortality

Over more than a decade, time-series epidemiologic studies have been conducted in many cities to evaluate the association between daily changes in concentrations of particulate matter in the air and daily counts of morbidity and mortality. The HEI-funded National Morbidity, Mortality, and Air Pollution Study (NMMAPS) was designed to address concerns about bias in the selection of cities in air pollution studies, to allow effects of PM to be estimated more precisely, and to explore heterogeneity of effects. NMMAPS investigators at Johns Hopkins University and Harvard University selected multiple cities only according to population size and availability of monitoring data for particulate matter less than 10  $\mu\text{m}$  in aerodynamic diameter (PM<sub>10</sub>).

Findings from NMMAPS support an association between short-term increases in the concentration of PM<sub>10</sub> and mortality in the 90 largest cities in the United States. The NMMAPS investigators reported similar findings for hospitalizations for cardiovascular and respiratory diseases and mortality among residents 65 years of age and older in a smaller group of cities with daily PM<sub>10</sub> monitoring. One of HEI's original objectives in funding NMMAPS was to contribute to an evaluation of coherence of epidemiologic findings by investigating whether associations of PM<sub>10</sub> with deaths and with hospitalizations were related within each city. The parallel time series of mortality and hospitalizations allowed the NMMAPS investigators to pursue this objective while examining the effect of PM concentrations on two different, but presumably related, health outcomes.

*Coherence* is one of many criteria that have been applied in epidemiologic and other types of studies to assess whether the relation between an exposure

and health outcomes is likely to be causal. As originally proposed, coherence implied that a causal basis for an association did not conflict with what is known about the natural history or biology of disease. In air pollution epidemiology, coherence has come to mean that causality implies similar responses to air pollution for a number of related health outcomes. Using this sense of coherence in support of causation would require that associations between short-term increases in air pollution concentrations and mortality be similar to those for hospitalizations, for example, as an indicator of morbidity. This is the sense of coherence that motivated HEI to support the work conducted in NMMAPS IV.

### APPROACH

Several questions could be addressed by the combined analysis of air pollution effects on mortality and on hospital admissions. This study pursues one: Is the underlying true effect per unit PM<sub>10</sub> on mortality (the mortality slope) of the same magnitude as the effect per unit PM<sub>10</sub> on hospitalizations (the hospitalization slope) in a given city? (In other words, are the true mortality slopes and the true hospitalization slopes associated? The true slope is the slope that would be found in the absence of sampling error, for example, if each city's time series were extremely long.)

The investigators conducted this study by using data from 10 cities with daily PM<sub>10</sub> monitoring and daily mortality and hospitalization data. They restricted analyses to deaths and hospitalizations due to cardiovascular diseases in residents 65 years of age and older. They used a standardized analytic approach, even though the previous NMMAPS

*Continued*

## Research Report 94 Part IV

used somewhat different approaches for the analyses of mortality and hospitalization in these cities. Any lack of correlation could not, then, be attributed to differences in analytic approach.

For each city, the investigators used methods they had developed earlier to evaluate the association between PM<sub>10</sub> concentration with mortality and with hospitalizations, separately. They then developed and applied a new method to estimate the correlation between the associations of PM<sub>10</sub> concentration with mortality and with hospitalization in each city, taking into account the cross-correlation between the mortality and hospitalization time series. In a second stage of analysis, the investigators applied previously developed Bayesian hierarchical methods to estimate the correlation between the associations of PM<sub>10</sub> concentration with mortality and with hospitalization across all cities while accounting for variability due to sampling error.

### RESULTS

There was little or no correlation between the time series of daily deaths and of daily hospitalizations within each city. The correlations between estimated effects (associations) of PM<sub>10</sub> on daily deaths and hospitalizations within each city were estimated to range from  $-0.05$  to  $0.34$ . The estimated between-city correlation between effect estimates for mortality and hospitalization was low ( $0.20$ ) but was estimated with a large degree of uncertainty (95% posterior interval  $-0.89, 0.98$ ).

For each city, Bayesian estimates that assumed no correlation between the mortality and hospitalization PM<sub>10</sub> effect estimates were similar to Bayesian estimates from a joint analysis in which these correlations were taken into account. The similarity of the results reflects weak correlation. The investigators attempted to use mortality data from a city for which no hospitalization data were available, in this case New York, in order to estimate a PM<sub>10</sub> effect on hospitalizations in this city. They found that including these additional data did not much

improve the precision of the estimated PM<sub>10</sub> effect on New York hospitalizations compared with using only mortality and hospitalization data from other cities to calculate this estimate.

### CONCLUSIONS

The Special Panel of the Health Review Committee that was convened to review results from NMMAPS research reached the following conclusions:

1. The main contribution of NMMAPS Part IV is methodologic.
2. The Investigators' Report describes, for the first time, a method for estimating correlation between effects of a covariate (PM<sub>10</sub>) on two parallel time series of counts of population health outcomes over multiple cities. The ability to conduct a joint inference on two separate but related outcomes aids quantification of the extent of coherence of the data.
3. The mortality and hospitalization estimates of effect were only weakly correlated, providing no support for coherence when assessed in this way. Because of the imprecise estimate of the correlation, however, the question of coherence in the context of observational studies could not be adequately assessed. The broader question of coherence within the complete body of work on PM<sub>10</sub> (including toxicologic and other experimental findings) is not addressed by these methods.
4. Although the methods could be used to predict hospitalization effect estimates from mortality effect estimates, their utility remains to be demonstrated, given the low correlation between the mortality and hospitalization effect estimates in this study.
5. The finding of low correlations between daily counts of deaths and daily counts of hospitalizations is noteworthy. It suggests that most mechanisms causing fluctuations in these counts over time—whether due to PM<sub>10</sub> or other factors—differ for the two outcomes.



# CONTENTS

## Research Report 94 Part IV

HEALTH  
EFFECTS  
INSTITUTE

### The National Morbidity, Mortality, and Air Pollution Study

#### Part IV: Hierarchical Bivariate Time-Series Models—A Combined Analysis of PM<sub>10</sub> Effects on Hospitalization and Mortality

Francesca Dominici, Antonella Zanobetti, Scott L Zeger, Joel Schwartz, and Jonathan M Samet

*Department of Biostatistics and Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts*

#### HEI STATEMENT

This Statement is a nontechnical summary of the Investigators' Report and the Health Review Committee's Commentary.

#### INVESTIGATORS' REPORT

When an HEI-funded study is completed, the investigators submit a final report. The Investigators' Report is first examined by three outside technical reviewers and a biostatistician. The report and the reviewers' comments are then evaluated by members of the HEI Health Review Committee, who had no role in selecting or managing the project. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, if necessary, revise the report.

Abstract . . . . .	1	Discussion and Conclusions . . . . .	10
Introduction . . . . .	1	Acknowledgments . . . . .	12
Specific Aims and Modeling Approach . . . . .	2	References . . . . .	12
Study Data . . . . .	3	About the Authors . . . . .	14
Statistical Methods and Data Analysis . . . . .	4	Other Publications Resulting from This Research . . . . .	15
Results . . . . .	6	Abbreviations and Other Terms . . . . .	15

#### COMMENTARY Special Panel of the Health Review Committee

The Commentary about the Investigators' Report is prepared by the HEI Health Review Committee and staff. Its purpose is to place the study into a broader scientific context, to point out its strengths and limitations, and to discuss remaining uncertainties and implications of the findings for public health.

Introduction . . . . .	17	Results . . . . .	23
Study Objectives . . . . .	18	NMMAPS IV and Coherence . . . . .	24
Analytic Methods . . . . .	18	Conclusions . . . . .	24
Results . . . . .	21	Acknowledgments . . . . .	25
Critique . . . . .	22	References . . . . .	25
Methods . . . . .	22		

#### RELATED HEI PUBLICATIONS

---

Publishing history: This document was posted as a preprint on [www.healtheffects.org](http://www.healtheffects.org) and then finalized for print.

Citation for whole document:

Dominici F, Zanobetti A, Zeger SL, Schwartz J, Samet JM. September 2005. The National Morbidity, Mortality, and Air Pollution Study, Part IV: Hierarchical Bivariate Time-Series Models—A Combined Analysis of PM<sub>10</sub> Effects on Hospitalization and Mortality. Research Report 94. Health Effects Institute, Boston MA.

When specifying a section of this report, cite it as a chapter of the whole document.

## The National Morbidity, Mortality, and Air Pollution Study

### Part IV. Hierarchical Bivariate Time-Series Models—A Combined Analysis of PM<sub>10</sub> Effects on Hospitalization and Mortality

Francesca Dominici, Antonella Zanobetti, Scott L Zeger, Joel Schwartz, and Jonathan M Samet

---

#### ABSTRACT

---

In this paper we describe our development of a hierarchical bivariate time-series model to characterize the relation between particulate matter less than 10  $\mu\text{m}$  in aerodynamic diameter (PM<sub>10</sub>\*) and both mortality and hospital admissions due to cardiovascular diseases. The model was applied to time-series data on mortality and hospital admissions for 10 metropolitan areas (henceforth called *cities*) in the United States from 1986 to 1993. We hypothesized that these time series are related through a shared relation with PM<sub>10</sub>.

In the first stage of the hierarchical model, we fit two seemingly unrelated Poisson regression models to estimate the log-relative rates of mortality and hospital admissions associated with exposure to PM<sub>10</sub> within each city. The sample covariance matrix of the estimated log-relative rates was obtained using a novel generalized estimating equation approach that takes into account the correlation between the mortality and hospital admission time series. In the second stage, we combined information across cities to estimate overall log-relative rates of mortality and hospital admission and variation in the rates across cities.

Using the combined information across the 10 cities, we found that a 10  $\mu\text{g}/\text{m}^3$  increase in mean PM<sub>10</sub> for the current day and the previous day is associated with a 0.26% increase in mortality for cardiovascular diseases (95% posterior interval  $-0.37, 0.65$ ) and with a 0.71% increase in

hospital admissions for cardiovascular diseases (95% posterior interval 0.35, 0.99). The log-relative rates of mortality and hospital admission had a similar degree of heterogeneity across cities: the posterior means of the between-city SDs of mortality and hospital admission PM<sub>10</sub> effects were 0.42 (95% posterior interval 0.05, 1.18) and 0.31 (95% posterior interval 0.10, 0.89), respectively. The city-specific log-relative rates of mortality and hospital admission were estimated to correlate very weakly, but the uncertainty in the correlation was substantial (posterior mean = 0.20, 95% posterior interval  $-0.89, 0.98$ ).

With the parameter estimates from the model, we could predict the log-relative rate of hospitalization for a city for which hospitalization data are unavailable, by using that city's estimated relative rate of mortality. We illustrate this prediction using New York as an example.

---

#### INTRODUCTION

---

The potential for high concentrations of air pollution to cause excess deaths and morbidity was firmly established in the mid 20th century by a series of well-documented episodes designated as air pollution disasters in the United States and Europe. By the mid 1990s, time-series studies with either mortality or morbidity data from single locations (Dockery et al 1993; Schwartz, 1994a; Bascom et al 1996; Pope 2000) showed that air pollution, even at much lower concentrations than existed during the earlier disasters, was associated with increased rates of mortality and morbidity in cities in the United States, Europe, and other developed countries (Pope and Dockery 1999). One key limitation of these studies is the use of data from a single location, or at most a few locations, that may not have represented broader geographic regions. A second limitation is that mortality and morbidity measures were considered separately.

The US National Morbidity, Mortality, and Air Pollution Study (NMMAPS) addressed these limitations by assembling and analyzing a national database that includes information about mortality, morbidity, weather, and air pollution for numerous metropolitan areas in the United States.

---

\*A list of abbreviations and other terms appears at the end of the Investigators' Report.

This Investigators' Report is one part of Health Effects Institute Research Report 94 Part IV, which also includes a Commentary by a Special Panel of the Health Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr Jonathan M Samet, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St, Suite 6041, Baltimore MD 21205-2179.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R82811201 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

The NMMAPS mortality analyses estimated associations between all-cause and cause-specific mortality and PM<sub>10</sub> levels for 90 cities in the United States (Samet et al 2000c; Dominici et al 2002). The NMMAPS morbidity analyses estimated associations between hospitalization in older persons and PM<sub>10</sub> levels for 14 cities in the United States (Samet et al 2000c; Schwartz 2000; Zanobetti et al 2000a). Methodologic approaches and substantive results of the separate mortality and morbidity analyses have been reported (Zeger et al 1999; Daniels et al 2000, 2004; Samet et al 2000a; Schwartz 2000; Schwartz and Zanobetti 2000; Zanobetti et al 2000b). These analyses showed that PM<sub>10</sub> concentrations were, on average, positively associated with mortality and morbidity outcomes across cities (Samet et al 2000c; Dominici et al 2003; Schwartz et al 2003).

Poisson time-series regression models (Liang and Zeger 1986; Zeger and Liang 1992; McNeney and Petkau 1994; Albert 1999; Fahrmeir and Tutz 2001) and generalized additive models (Hastie and Tibshirani 1990) have been widely used to analyze univariate time-series data of air pollution and health in selected locations (Dockery and Pope 1994; Schwartz 1995; Bascom et al 1996a,b; Korrick et al 1998). Critics of single-site studies have questioned the choice of particular cities and have asked whether models were selected that gave estimates of effect that were biased upwards (Lipfert and Wyzga 1993; Li and Roth 1995). These criticisms have been addressed by multisite studies in which site-specific data on air pollution and health were assembled under a common framework (Katsouyanni et al 1997; Samet et al 2000a; Hwang and Chan 2002).

Hierarchical models (DuMouchel and Harris 1983; DuMouchel 1990; Breslow and Clayton 1993; Carlin and Louis 1996) are suitable for analyzing univariate time-series data from multiple locations (Dominici et al 2000, 2002; Hwang and Chan 2002). Compared with analyses of data from a single site, pooled analyses can be more informative about whether an association exists after controlling for possible confounders. In addition, pooled analyses can estimate parameters at a specific site, which borrow strength from all other locations.

Hierarchical univariate time-series models used to estimate associations between air pollution and health are focused mainly on estimating the overall log-relative rate of mortality or morbidity associated with PM<sub>10</sub> levels and their heterogeneity across locations. Such analyses have not explored whether cities in which PM<sub>10</sub> had greater or lesser effects on morbidity also tended to show a similar pattern of PM<sub>10</sub> effects on mortality.

The purpose of this study was to explore across-city correlation between the underlying true effects of PM<sub>10</sub> on mortality and on hospitalization. We expected a correlation

to exist because the same city-specific processes modify the PM<sub>10</sub> effects on both outcomes. For example, in a community in which buildings are well ventilated, we might expect the slope of the personal exposure–ambient PM<sub>10</sub> concentration relation to be higher than average. The greater rate of air exchange would make PM<sub>10</sub> concentrations closer to those outdoors. All other things being equal, that would result in a higher coefficient relating ambient PM<sub>10</sub> concentration to health for either outcome, because a 1 µg/m<sup>3</sup> increment in ambient concentration represents a larger than average increment in exposure. Both outcomes would be equally affected by this process. Coefficients for both hospital admissions and mortality would be expected to vary with prevalence rates of conditions and diseases related to susceptibility to PM<sub>10</sub>.

---

## SPECIFIC AIMS AND MODELING APPROACH

---

An overall goal of this study was to extend the prior NMMAPS analyses by jointly analyzing data on mortality and hospital admission within the same city. We developed methods for joint analyses of hospital admission and mortality data and applied them to data for 10 cities, for which nearly daily PM<sub>10</sub> monitoring, mortality, and hospitalization data were available.

Our modeling approach extends previous work in two directions. First, within each city we extended Poisson regression approaches for univariate time series to bivariate time series and we estimated the log-relative rates of mortality and hospital admission by taking into account the correlation between the daily mortality and hospital admission time series. Second, we extended this bivariate time-series model in a hierarchical fashion, by combining relative rates of mortality and hospital admission across cities in order to characterize their relation.

More specifically, in the first stage, we fit two *seemingly unrelated* Poisson regression models to estimate the relative rates of mortality (M) and hospital admission (H) due to cardiovascular diseases associated with exposure to PM<sub>10</sub> ( $\hat{\beta}_M^c, \hat{\beta}_H^c$ ) within each city  $c$ . We defined these two Poisson regression models as seemingly unrelated for the following two reasons. First, we estimated  $\hat{\beta}_M^c$  and  $\hat{\beta}_H^c$  under the working assumption that the daily mortality and hospitalization time series are independent. Secondly, to take into account the joint correlation function for the bivariate mortality and hospital admission time series, we estimated the sample covariance between  $\hat{\beta}_M^c$  and  $\hat{\beta}_H^c$  by using a novel generalized estimating equations approach (Zeger et al 1988; Dominici et al 2004b).

In the second stage, we assumed that the vector of true log-relative rates of mortality and hospital admission ( $\beta_M^c$  and  $\beta_H^c$ ) has a bivariate normal distribution with unknown means ( $\alpha_M$  and  $\alpha_H$ ) and an unknown covariance matrix ( $\Sigma$ ). We then estimated these unknowns using simulation-based techniques (Gilks et al 1996; Everson and Morris 2000).

Although the sample covariance between  $\hat{\beta}_M^c$  and  $\hat{\beta}_H^c$  and the correlation between the mortality and hospital admission time series within each city are of interest, we focused on the parameters from the second stage of the hierarchical model (this choice is justified in the Introduction and Discussion).

Our hierarchical bivariate time-series model can be used to predict the log-relative rates of mortality and hospital admission for cities other than the 10 included in the joint analysis. For example, consider New York, for which we have mortality data but not hospital admission data. We approximated the posterior predictive distribution of the log-relative rate of hospital admissions for New York ( $\beta_H^{NY}$ ) conditional on the New York mortality data and the mortality and hospital admission data for the other 10 cities. In addition, we estimated the reductions in the posterior variances of  $\beta_M^c$  and  $\beta_H^c$  obtained by including the time-series data at city  $c$  relative to not including this information. We report these reductions in posterior variances for all 10 cities and for New York.

## STUDY DATA

For this study, the general observation period for the air pollution, mortality, hospital admission, and meteorologic data for the 10 US cities was 1986 to 1993; the city-specific period varied depending on data availability (Table 1). The air pollution data were obtained from the Aerometric Information Retrieval System (AIRS) database maintained by the US Environmental Protection Agency. The daily time series of  $PM_{10}$  used for these analyses were the same as those used for the NMMAPS morbidity analysis (Samet et al 2000c; Schwartz et al 2003). Daily counts of mortality from all causes for residents 65 years of age and older, aggregated at the county level, were obtained from the US National Center for Health Statistics (NCHS). Daily counts of hospital admissions for residents 65 years of age and older were extracted from the files of the US Health Care Financing Administration (HCFA, renamed the Centers for Medicare & Medicaid Services in July 2001). Hourly temperatures, barometric pressures, and relative humidities for each site within each city were obtained from the Earth Info CD-ROM database ([www.earthinfo.com](http://www.earthinfo.com)). More details about the data are given elsewhere (Samet et al 2000b,c).

For this report, we focused on cardiovascular diseases because prior research has suggested these are the diseases most strongly associated with variation in air quality (Dockery et al 1993; Samet et al 2000b). We chose the

**Table 1.**  $PM_{10}$  Monitoring Period and Pollutant and Outcome Data for This Study<sup>a</sup>

City and State <sup>b</sup>	PM <sub>10</sub> Monitoring		# Days PM <sub>10</sub> Data Available	24-Hour Mean PM <sub>10</sub> Concentration ( $\mu\text{g}/\text{m}^3$ )	Daily Mean #	
	Start Date	End Date			Hospital Admissions <sup>c</sup>	Deaths <sup>c</sup>
Birmingham AL (1)	3/1/87	12/31/93	2485	34.8	24	6
Canton OH (2)	1/1/88	12/31/93	1750	28.4	10	3
Chicago IL (3)	1/1/88	12/31/93	2058	36.3	114	48
Colorado Springs CO (4)	7/1/87	12/31/93	2310	27.5	3	2
Detroit MI (5)	4/1/86	12/31/93	2517	36.7	53	23
Minneapolis/St Paul MN (6)	2/1/87	12/31/93	2488	28.1	22	10
New Haven CT (7)	1/1/88	12/31/91	1450	28.6	19	8
Pittsburgh PA (8)	1/1/86	12/31/93	2918	36.0	51	16
Seattle WA (9)	1/1/86	12/31/93	2913	32.2	20	9
Spokane WA (10)	1/1/86	12/31/93	2778	42.9	6	3
New York NY (example)	1/1/87	12/31/94	489	28.8	NA <sup>d</sup>	108

<sup>a</sup> From the database we also used 24-hour mean temperature, barometric pressure, and relative humidity data.

<sup>b</sup> Identification numbers given in parentheses are used in Figure 1.

<sup>c</sup> Due to cardiovascular diseases.

<sup>d</sup> Not available.

10 largest cities from the 14 originally included in the NMMAPS hospitalization analysis for which daily time series of mortality were available for the same time period as for the daily time series of hospital admissions (Table 1; Samet et al 2000c; Schwartz et al 2003).

---

## STATISTICAL METHODS AND DATA ANALYSIS

---

We developed a two-stage hierarchical model for combining log-relative rates of mortality and hospital admissions across cities, taking into account the joint correlation function for the bivariate mortality and hospital admission time series when estimating the covariance of the two log-relative rates. (See also Rochon [1996] for statistical methods for analyzing bivariate repeated outcomes.) The goals of our analysis were to estimate (1) overall log-relative rates of mortality and hospital admissions due to cardiovascular diseases associated with exposure to PM<sub>10</sub>; (2) heterogeneity of the two log-relative rates across cities; (3) correlation between the two log-relative rates across cities; and (4) the log-relative rate of hospital admissions for a city other than the 10 sampled, using that city's mortality time-series data. The two stages of the hierarchical model are described below.

In the first stage, for within each city, we fit two seemingly unrelated log-linear regressions to the mortality (M) and hospital admissions (H) data. We assumed

$$\begin{aligned}
 E(y_t^M) &= \mu_t^M, \text{ var}(y_t^M) = \phi^M \mu_t^M \\
 \log \mu_t^M &= X_t' \theta_M \\
 E(y_t^H) &= \mu_t^H, \text{ var}(y_t^H) = \phi^H \mu_t^H \\
 \log \mu_t^H &= X_t' \theta_H
 \end{aligned} \tag{1}$$

where  $y_t^M$  and  $y_t^H$  are the mortality and hospital admissions daily time series and  $X_t$  is the design matrix, including mean lag 0 and lag 1 PM<sub>10</sub> daily time series and possible confounding factors for the mortality and hospitalization data (such as long-term time trends, seasonality, and weather; Samet et al 1995, 1997, 2000a; Kelsall et al 1997; Dominici et al 2000).  $\theta$  is the corresponding vector of regression coefficients; var denotes variance and  $E$  denotes the expected value;  $\phi$  is the overdispersion parameter; and  $\mu$  is the expected value of the random variable  $y$ .

In this application, we specified model (1) as an overdispersed Poisson regression model with a linear term for the mean PM<sub>10</sub> concentration on days 0 and 1. We also specified smooth functions (natural cubic splines) of calendar time, temperature, barometric pressure, and relative humidity to adjust for time-varying confounding factors (such as time

trend, seasonality, and weather). The rationales for inclusion of all variables in the model are listed in Table 2.

Thus, the full vector of regression coefficients denoted by  $\theta_M^c$  (or  $\theta_H^c$ ) for city  $c$  in model (1) can be decomposed as  $[\beta_M^c, \eta_M^c]$  (or  $[\beta_H^c, \eta_H^c]$ ) where  $\beta_M^c$  (or  $\beta_H^c$ ) is the log-relative rate of mortality (or hospital admission) for increases in PM<sub>10</sub> level and  $\eta_M^c$  (or  $\eta_H^c$ ) is the vector of nuisance parameters corresponding to the confounding factors listed in Table 2. Finally, the parameters  $\phi_M$  and  $\phi_H$  are overdispersion parameters.

Modeling strategies to reduce confounding bias in air pollution effect estimates are among the most discussed statistical issues in time-series analyses of air pollution and health. In particular, choosing the number of degrees of freedom ( $df$ ) to include in the analytic model for smooth functions of time and temperature are critical, because they determine the residual temporal variability in daily deaths and pollution levels that is used to estimate the pollution coefficient. As a baseline in the model, we used 4  $df$  per year to adjust for time trend and seasonality and 3  $df$  to adjust for temperature, barometric pressure, and relative humidity. These choices were made on the basis of our previously published results (Samet et al 2000b,c) and results of recent revised analyses and sensitivity analyses (Dominici et al 2003).

We then explored the sensitivity of the overall log-relative rates to the adjustment for confounding factors. We estimated overall log-relative rates of mortality ( $\alpha_M$ ) and hospital admission ( $\alpha_H$ ) in correspondence to five alternative adjustment scenarios. The five scenarios were defined by multiplying the number of degrees of freedom of the smooth functions of time, temperature, barometric pressure, and relative humidity (natural cubic splines as defined in Table 2) by a calibration parameter  $\delta$  that assumes values 1/3, 1/2, 1, 2, and 3. Compared with our baseline model,  $\delta = 1$ ,  $\delta = 1/3$ , and  $\delta = 3$  represent less and more dramatic adjustments for time trend, seasonality, and weather. The rationale behind the choice of the five adjustment scenarios tested, and more extensive results of the sensitivity of the overall log-relative rate of mortality for the largest 90 cities, are discussed and summarized elsewhere (Dominici et al 2004a).

We estimated the log-relative rate parameters ( $\hat{\beta}_M^c, \hat{\beta}_H^c$ ) for city  $c$  under the working assumption that the daily mortality and hospitalization time series are independent. Hence, two separate log-linear regressions were estimated by maximum likelihood. We consider this approach to be sensible because we focused on the association between the log-relative rates rather than the association between the daily counts. However, correlation between the two series of counts introduces correlation in the estimated

**Table 2.** Model Specification<sup>a</sup> for Estimating City-Specific Log-Relative Rates Associated with Current Day (lag 0) and Previous Day (lag 1) PM<sub>10</sub> Concentrations

Predictor	Primary Reasons for Inclusion
Average PM <sub>10</sub> at lag 0 and lag 1 (linear term)	To estimate log–log relative rates of mortality associated with short-term increase in PM <sub>10</sub> concentrations
Indicator variables for day of the week (linear terms)	To allow different baseline log mortality rate for each day of the week
Smooth functions of time (4 <i>df</i> × year)	To adjust for long-term trend and seasonality
Smooth functions of temperature at lag 0 and lag 1 (3 <i>df</i> )	To control for known effects of weather on mortality
Smooth functions of barometric pressure and relative humidity (3 <i>df</i> )	To control for known effects of humidity on mortality

<sup>a</sup> We used an overdispersed Poisson regression model that included indicator variables for day of the week and specified smooth functions of time, temperature, barometric pressure, and relative humidity as natural cubic splines.

log-relative rates,  $\hat{\beta}_M^c$  and  $\hat{\beta}_H^c$ , for a given city. We used generalized estimating equations to estimate the sample covariance matrix  $V^c$ , the overdispersion parameters  $\phi_M$  and  $\phi_H$ , and the within-city correlation

$$v_{MH}^c / \sqrt{v_M^c v_H^c}$$

(where  $v_M^c$  and  $v_H^c$  are the sample variances of  $y_t^H$  and  $y_t^M$ , respectively, and  $v_{MH}^c$  denotes the sample covariance between  $y_t^H$  and  $y_t^M$ ). To do so, we applied equation A.4, detailed in Dominici and colleagues (2004b), for a lag of 14 days. This choice is based on the assumption that the mortality and hospital admission time series are likely to be uncorrelated at lag 14. The estimated  $V^c$  matrices were not sensitive to lag choices larger than 14.

Because the time series were relatively long (PM<sub>10</sub> concentration data were available for  $\geq 1450$  days), the estimated mortality and hospital admission log-relative rates are approximately bivariate normal (denoted by  $N_2$ ):

$$\begin{bmatrix} \hat{\beta}_M^c \\ \hat{\beta}_H^c \end{bmatrix} \sim N_2 \left( \begin{bmatrix} \beta_M^c \\ \beta_H^c \end{bmatrix}, V^c \right) \text{ where } V^c = \begin{bmatrix} v_M^c & v_{MH}^c \\ v_{MH}^c & v_H^c \end{bmatrix}. \quad (2)$$

Dominici and colleagues (2000) have shown that this approximation to the likelihood is accurate for estimating overall log-relative rates.

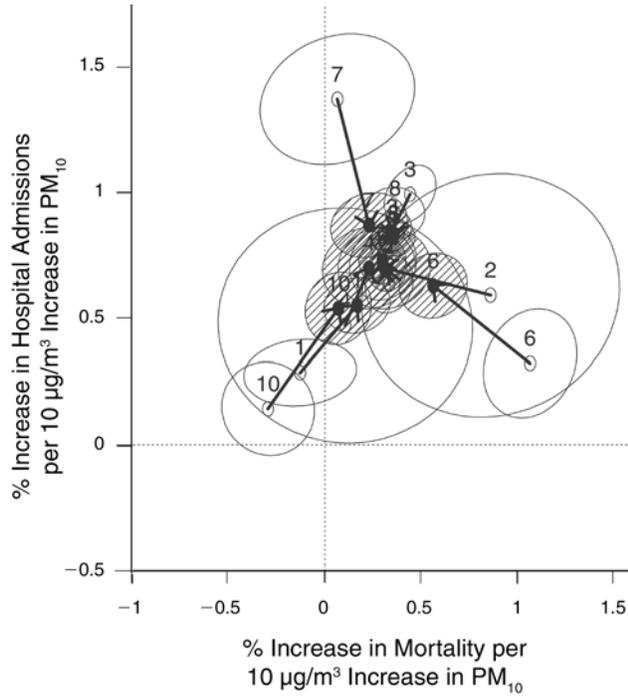
The second stage of the hierarchical model describes variation among the true log-relative rates  $\beta_M^c$  and  $\beta_H^c$  across cities. We assumed:

$$\begin{bmatrix} \beta_M^c \\ \beta_H^c \end{bmatrix} \sim N_2 \left( \begin{bmatrix} \alpha_M \\ \alpha_H \end{bmatrix}, \Sigma \right) \text{ where } \Sigma = \begin{bmatrix} \sigma_M^2 & \sigma_{MH} \\ \sigma_{MH} & \sigma_H^2 \end{bmatrix}. \quad (3)$$

Here  $\alpha_M$  and  $\alpha_H$  denote the overall log-relative rates of mortality and hospital admissions associated with exposure to PM<sub>10</sub>;  $\sigma_M^2$  and  $\sigma_H^2$  are the variances of  $\beta_M^c$  and  $\beta_H^c$ ; and  $\rho_{MH} = \sigma_{MH} / \sigma_M \sigma_H$  denotes the correlation across cities between  $\beta_M^c$  and  $\beta_H^c$ . Larger values of  $\rho$  indicate that cities with higher log-relative rates of mortality per 10- $\mu\text{g}/\text{m}^3$  increase in PM<sub>10</sub> concentration are also more likely to have higher log-relative rates of hospital admissions per 10- $\mu\text{g}/\text{m}^3$  increase in PM<sub>10</sub> concentration.

Specification of this Bayesian hierarchical model was completed by assigning prior distributions for the parameters. For the mean parameters ( $\alpha_M$  and  $\alpha_H$ ), we assumed vague normal priors having a mean of 0 with large variances. Under the two-stage multivariate normal model (3), a natural choice for the prior distribution on the covariance matrix is the conjugate prior inverse Wishart distribution. Although the inverse Wishart distribution is mathematically convenient for implementation of simulation-based techniques (Gilks et al 1996), it is not flexible enough to elicit noninformative priors on the variances and on the correlation coefficient (Daniels 1999; Daniels and Kass 1999). Instead of assuming the conjugate prior for the entire covariance matrix, we assumed that the two variance components  $\sigma_M^2$  and  $\sigma_H^2$  have, a priori, a half-normal distribution on  $(0, \infty)$  with a mean of 0 and a large variance (here, chosen to be 10) and that the correlation coefficient  $\rho$  has, a priori, a uniform distribution in  $[-1, 1]$ . This prior specification is denoted as our *baseline prior on  $\Sigma$*  throughout the report.

We applied the methods described in Dominici and colleagues (2004b) to estimate the sample covariance matrix  $V^c$  of the log-relative rate estimates,  $\hat{\beta}_M^c$  and  $\hat{\beta}_H^c$ , within each city. Posterior distributions of all parameters of interest were approximated by simulation-based techniques (Gelman et al 1995).



**Figure 1.** 10% Highest likelihood density regions (encircled by solid lines) and 10% highest posterior density regions (hatched) of the log-relative rates of mortality and hospital admissions for cardiovascular diseases. Maximum likelihood estimates and Bayesian estimates of the log-relative rates are connected with arrows. Number labels refer to cities (see Table 1). The Bayesian estimates were obtained under our baseline prior for the covariance matrix  $\Sigma$ . (○) Maximum likelihood estimate; (●) Bayesian estimate.

## RESULTS

Figure 1 shows the 10% highest likelihood density regions and 10% highest posterior regions of the log-relative rates of mortality and hospitalization for each of the 10 cities. The Bayesian estimates were obtained under our baseline prior for  $\Sigma$  (ie, half-normal variances and uniform correlation). The shapes of the likelihood density regions indicate that the within-city estimated statistical correlations between the estimated effects of PM<sub>10</sub> on mortality and on hospitalizations ( $v_{HM}^c/\sqrt{v_{HH}^c v_{MM}^c}$ ) are small. The statistical correlations between the estimated effects of PM<sub>10</sub> on mortality and on hospitalizations ranged from  $-0.05$  in Spokane to  $0.34$  in Chicago. The low statistical correlations between estimated effects of PM<sub>10</sub> on mortality and on hospitalizations indicate that the mortality and hospital admissions time series were only weakly correlated. City-specific estimates of the overdispersion parameters ( $\hat{\phi}_M, \hat{\phi}_H$ ) and of the within-city correlations between the estimated effects of PM<sub>10</sub> on mortality and on hospitalizations are listed in Table 3, along with other estimates.

Maximum likelihood and Bayesian estimates of the log-relative rates for mortality and for hospital admissions for the 10 cities (with their 95% confidence intervals and 95% posterior regions, respectively) are also given in Table 4. Results are also reported under a separate analysis that assumes that mortality (or hospital admission) data do not provide any information about the log-relative rates of hospital admissions (or mortality) (eg,  $\rho = 0$ ). The Bayesian estimates of the log-relative rates from the combined analysis are very similar to those from the separate analysis, suggesting that  $\rho$  is very small or poorly estimated or both.

**Table 3.** Estimates of Overdispersion Parameters and Corresponding Within-City Pearson Correlation Coefficients

City	$\hat{\phi}_M^a$	$\hat{\phi}_H^a$	$(v_{MH}^c/\sqrt{v_{MM}^c v_{HH}^c})^b$	$\text{cor}(y_t^{cM}, y_t^{cH})^c$	$\text{cor}(r_t^{cM}, r_t^{cH})^d$	$\text{cor}(r_t^{cM}, r_{t-1}^{cH})^e$	$\text{cor}(r_t^{cM}, r_{t-2}^{cH})^f$
Birmingham	0.84	0.93	0.11	0.03	0.03	0.00	-0.01
Canton	0.82	0.96	0.07	0.04	0.01	0.03	0.02
Chicago	0.91	1.21	0.34	0.08	0.04	0.04	0.04
Colorado Springs	0.89	0.89	-0.02	0.00	0.02	-0.01	0.00
Detroit	0.83	1.01	0.14	0.00	-0.01	-0.01	0.01
Minneapolis/St Paul	0.82	0.94	0.18	-0.01	0.02	0.05	-0.02
New Haven	0.84	0.89	0.17	0.03	0.02	-0.04	-0.02
Pittsburgh	0.86	1.02	-0.02	0.00	-0.01	-0.01	0.01
Seattle	0.85	0.97	0.10	0.02	0.03	0.02	0.00
Spokane	0.88	0.87	-0.05	0.02	0.01	0.00	0.02

<sup>a</sup> City-specific overdispersion parameters for mortality (M) and hospital admissions (H).

<sup>b</sup> Within-city correlation between city-specific estimated slopes at lag 0.

<sup>c</sup> Within-city correlation between mortality and hospitalization time series at lag 0.

<sup>d</sup> Within-city correlation between residual mortality at lag 0 and residual hospitalization at lag 0.

<sup>e</sup> Within-city correlation between residual mortality at lag 0 and hospitalization time series at lag 1.

<sup>f</sup> Within-city correlation between residual mortality at lag 0 and hospitalization time series at lag 2.

To further quantify the statistical power for estimating  $\rho$ , we conducted a test of the heterogeneity of city-specific relative rate estimates. In the absence of heterogeneity,  $\rho$  is not identifiable. For the mortality and hospital admission relative rate estimates, we found that the  $\chi^2$  test statistics equal 15 and 85, respectively. Compared with the 5% critical level of a  $\chi^2$  distribution with 9 *df* (which is equal to 16.91), we rejected the hypothesis of heterogeneity for the mortality relative rate estimates, but we accepted it for the hospitalization relative rate estimates.

Because of the small number of cities, inferences about the degree of heterogeneity in pollution effects among cities are likely to be sensitive to the prior assumptions about  $\Sigma$ . Our strategy for investigating the effect of the prior distribution on our results was based on inspecting the posterior distributions of the parameters of interest under the following prior distributions for  $\Sigma$ :

- Uniform prior on  $\Sigma$  [eg, uniform prior on all entries in  $\Sigma$  with the condition that  $\Sigma$  is positive definite];
- Jeffrey prior on  $\Sigma$  [eg,  $p(\Sigma) \propto |\Sigma|^{-1}$ ]; and
- Uniform prior on the shrinkage matrix  $B_0$ , where

$$B_0 = V_0^{1/2} (V_0 + \Sigma)^{-1} V_0^{1/2}, \quad V_0 = \frac{1}{10} \sum_c^{10} V^c$$

[eg, uniform prior on all entries in matrix  $B$ ].

Additional details about these prior distributions, including the definitions of the prior densities and software implementations, are in Everson and Morris (2000).

The posterior distributions of the overall log-relative rates and of the between-city standard deviations under our baseline prior and under the alternative noninformative prior distributions are shown in Figure 2. The estimated overall relative rate of hospital admissions associated with  $\text{PM}_{10}$  expressed as percentage increase in mortality per 10  $\mu\text{g}/\text{m}^3$   $\text{PM}_{10}$  increase was 0.71% (95% posterior interval 0.35,0.99). The estimated overall log-relative rate of mortality was 0.26% per 10  $\mu\text{g}/\text{m}^3$   $\text{PM}_{10}$  (95% posterior interval  $-0.37,0.65$ ). Posterior distributions of the between-city standard deviations indicate the degree of heterogeneity of the mortality and hospital admission log-relative rates across cities. The distribution for mortality is similar to that for hospital admission, with posterior means of 0.42 for  $\sigma_M$  (95% posterior interval 0.05,1.18) and 0.31 for  $\sigma_H$  (95% posterior interval 0.10,0.89).

The posterior distributions of the between-city correlation coefficient  $\rho = \sigma_{MH}/\sigma_M\sigma_H$  are shown in Figure 3. The log-relative rates of mortality and hospital admission are weakly correlated and this correlation has great statistical uncertainty. The posterior mean of  $\rho$  between the true  $\beta_M^c$  and  $\beta_H^c$  was 0.20 (95% posterior interval  $-0.89,0.98$ ). Assuming a uniform prior on  $\Sigma$  led to larger posterior

**Table 4.** Maximum Likelihood and Bayesian Estimates of Log-Relative Rates of Mortality and Hospital Admissions for Cardiovascular Diseases Associated with a 10  $\mu\text{g}/\text{m}^3$  Increase in  $\text{PM}_{10}$  Concentration

City	Estimated Log-Relative Rates of Mortality			Estimated Log-Relative Rates of Hospital Admissions		
	Maximum Likelihood <sup>a</sup>	Bayes <sup>b</sup>		Maximum Likelihood <sup>a</sup>	Bayes <sup>b</sup>	
		Combined	Separate <sup>c</sup>		Combined	Separate <sup>c</sup>
Birmingham	-0.13(-1.38,1.12)	0.17(-0.57,0.90)	0.20(-0.52,0.92)	0.28(-0.28,0.84)	0.55(0.08,1.02)	0.48(-0.01,0.96)
Canton	0.86(-2.00,3.72)	0.31(-0.72,1.34)	0.32(-0.58,1.23)	0.59(-1.46,2.64)	0.70(0.00,1.39)	0.67(-0.13,1.47)
Chicago	0.44(-0.13,1.02)	0.35(-0.09,0.79)	0.36(-0.07,0.78)	0.99(0.50,1.48)	0.84(0.49,1.19)	0.85(0.45,1.25)
Colorado Springs	0.11(-2.71,2.92)	0.23(-0.80,1.26)	0.27(-0.58,1.13)	0.47(-1.51,2.45)	0.70(0.01,1.39)	0.67(-0.13,1.46)
Detroit	0.33(-0.20,0.87)	0.33(-0.10,0.75)	0.33(-0.09,0.74)	0.63(0.15,1.11)	0.69(0.33,1.04)	0.66(0.28,1.04)
Minneapolis/ St Paul	1.07(0.03,2.10)	0.56(-0.19,1.32)	0.56(-0.18,1.30)	0.32(-0.60,1.24)	0.63(0.08,1.17)	0.54(-0.07,1.15)
New Haven	0.07(-1.65,1.78)	0.23(-0.62,1.08)	0.23(-0.57,1.02)	1.36(0.26,2.47)	0.87(0.32,1.41)	0.89(0.18,1.60)
Pittsburgh	0.36(-0.30,1.03)	0.36(-0.12,0.83)	0.33(-0.16,0.82)	0.91(0.48,1.35)	0.82(0.50,1.13)	0.84(0.47,1.21)
Seattle	0.30(-0.44,1.04)	0.30(-0.22,0.82)	0.28(-0.23,0.80)	0.71(0.10,1.33)	0.73(0.33,1.12)	0.70(0.23,1.17)
Spokane	-0.29(-1.32,0.73)	0.07(-0.66,0.80)	0.09(-0.63,0.81)	0.14(-0.64,0.93)	0.54(-0.07,1.15)	0.47(-0.13,1.06)
New York	0.70(-0.18,1.58)	0.52(-0.10,1.14)	0.46(-0.19,1.12)	— <sup>d</sup>	0.61(-0.33,1.55)	— <sup>d</sup>
Overall	— <sup>d</sup>	0.26(-0.37,0.65)	0.28(-0.12,0.63)	— <sup>d</sup>	0.71(0.35,0.99)	0.69(0.33,1.06)

<sup>a</sup> 95% confidence intervals given in parentheses.

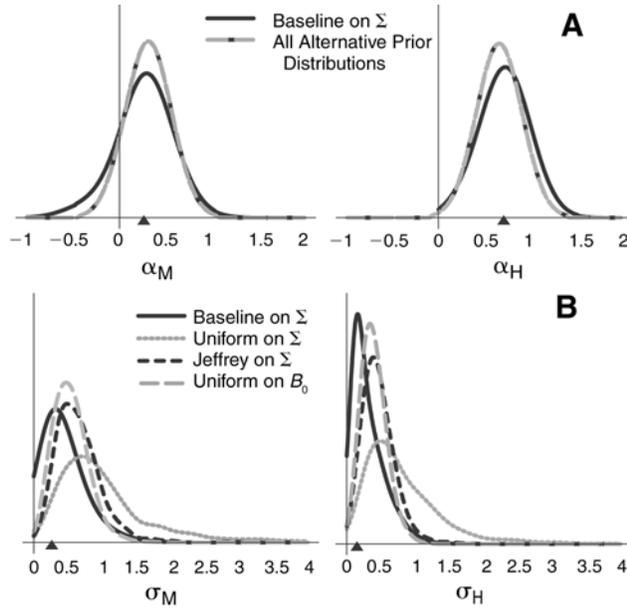
<sup>b</sup> Posterior means of log-relative rates, with 95% posterior regions given in parentheses.

<sup>c</sup> Results are from a model that assumes  $\rho = 0$ .

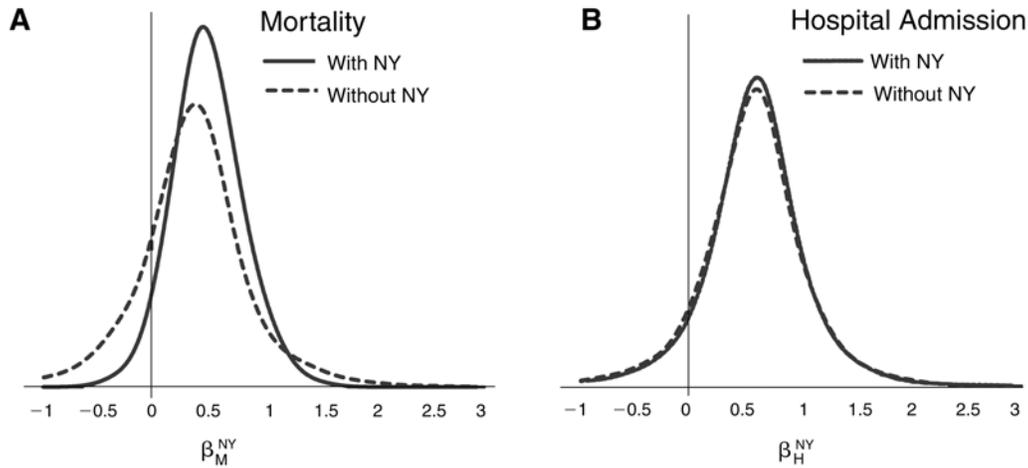
<sup>d</sup> Not applicable.

means and variances of  $\sigma_M^2$  and  $\sigma_H^2$  and to larger posterior variances of  $\alpha_M$  and  $\alpha_H$  but had little effect on  $\rho$ . The Jeffrey prior on  $\Sigma$  and the uniform prior on  $B_0$  gave posterior inferences nearly identical to those obtained with our baseline prior on  $\Sigma$  (half-normal variances and uniform correlation).

Generally, mortality time-series are more readily assembled from publicly available databases than are hospital admission time series. Therefore, predicting the log-relative rate of hospital admissions for a city (other than the 10 sampled) for which mortality but not hospital admission



**Figure 2.** Marginal posterior distributions (A) of overall log-relative rates of mortality ( $\alpha_M$ ) and hospital admissions ( $\alpha_H$ ) and (B) of standard deviations ( $\sigma_M$ ) and ( $\sigma_H$ ).  $\blacktriangle$  indicates posterior mean. Results for all alternative prior distributions (uniform on  $\Sigma$ , Jeffrey on  $\Sigma$ , and uniform on  $B_0$ ) in A were nearly identical.



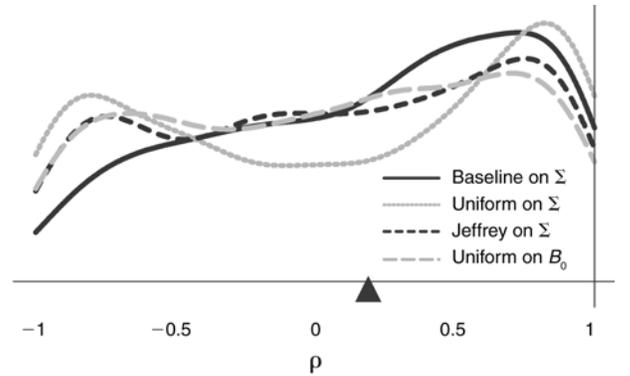
**Figure 4.** Distributions with and without New York mortality data. **A:** Marginal posterior distribution of  $\beta_M^{NY}$  including New York mortality data and posterior predictive distribution of  $\beta_M^{NY}$  not including New York mortality data. **B:** Posterior predictive distribution of  $\beta_H^{NY}$  including New York mortality data and posterior predictive distribution of  $\beta_H^{NY}$  not including New York mortality data.

data are available is desirable. We considered New York as an example. Using the model, we were also able to estimate reductions in the posterior variances of the log-relative rates of mortality and hospital admissions in New York ( $\hat{\beta}_M^{NY}$  and  $\hat{\beta}_H^{NY}$ ) and compare the values with and without use of the New York mortality time series.

Figure 4A shows the marginal posterior distribution of the mortality log-relative rate  $\beta_M^{NY}$  including the New York mortality data and the posterior predictive distribution of  $\beta_M^{NY}$  not including the New York mortality data. The marginal posterior distribution of  $\beta_M^{NY}$  (including mortality data for New York) was obtained by sampling from a univariate normal distribution with mean

$$\left( \frac{1}{\sigma_M^{2(j)}} + \frac{1}{\nu_M^{NY}} \right) \left( \frac{\alpha_M^{(j)}}{\sigma_M^{2(j)}} + \frac{\hat{\beta}_M^{NY}}{\nu_M^{NY}} \right)$$

where  $\alpha^{(j)}$  and  $\Sigma^{(j)}$  are the samples from the marginal posterior distribution of  $\alpha$  and  $\Sigma$  and  $\hat{\beta}_M^{NY}$  and  $\nu_M^{NY}$  are the maximum



**Figure 3.** Marginal posterior distributions of the correlation coefficient  $\rho$  ( $\rho = \sigma_{MH}/\sigma_M\sigma_H$ ).  $\blacktriangle$  indicates posterior mean.

likelihood estimate of the log-relative rate of mortality and the sample variance, respectively, for New York. The predictive distribution of  $(\beta_M^{NY}, \beta_H^{NY})$  (not including New York mortality data) was obtained by sampling from the bivariate normal distribution  $N(\alpha^{(j)}, \Sigma^{(j)})$ . The predictive distribution of  $\beta_H^{NY}$  (including the New York mortality data) was obtained by sampling from a normal distribution with mean

$$\alpha_H^{(j)} + \frac{\sigma_{MH}^{2(j)}}{\sigma_M^{2(j)}} (\beta_M^{NY(j)} - \alpha_M^{(j)})$$

where  $\beta_M^{NY(j)}$  is a sample from the marginal posterior distribution of  $\beta_M^{NY}$ . As expected, use of data from New York improves the estimate of  $\beta_M^{NY}$ , with a reduction in posterior variance of 65% (see also Table 5).

Figure 4B shows the posterior predictive distribution of the hospital admission log-relative rate  $\beta_H^{NY}$  including the New York mortality data and the posterior predictive distribution of  $\beta_H^{NY}$  not including the New York mortality data. The reduction in the posterior variance of  $\beta_H^{NY}$  obtained by including the mortality data in New York is 15%, much smaller than that of  $\beta_M^{NY}$  (see also Table 5). The relatively small gain in precision of the Bayesian estimate of  $\beta_H^{NY}$  obtained by using the New York mortality data versus not using it is due to the great imprecision in estimating the correlation coefficient  $\rho$ .

Table 5 shows reductions in posterior variances of  $\beta_M^c$  and  $\beta_H^c$  obtained by including time-series data for city  $c$  compared with not including the same data. Percentage reductions in posterior variances of the log-relative rates of hospital admission and mortality were slightly greater

**Table 5.** Reductions (%) in Posterior Variances of Log-Relative Rates of Mortality ( $\beta_M^c$ ) and Hospital Admissions ( $\beta_H^c$ )<sup>a</sup>

City	Combined Analysis		Separate Analysis	
	Mortality	Hospital Admissions	Mortality	Hospital Admissions
	$1 - \left( \frac{\text{var}(\beta_M^c   \text{data})}{\text{var}(\beta_M^c   \text{data}^{-c})} \right)$	$1 - \left( \frac{\text{var}(\beta_H^c   \text{data})}{\text{var}(\beta_H^c   \text{data}^{-c})} \right)$	$1 - \left( \frac{\text{var}(\beta_M^c   \text{data}_M)}{\text{var}(\beta_M^c   \text{data}_M^{-c})} \right)$	$1 - \left( \frac{\text{var}(\beta_H^c   \text{data}_H)}{\text{var}(\beta_H^c   \text{data}_H^{-c})} \right)$
Birmingham	59	68	52	72
Canton	23	26	16	22
Chicago	86	83	81	81
Colorado Springs	34	26	24	31
Detroit	87	81	82	85
Minneapolis/ St Paul	62	60	47	58
New Haven	51	56	45	48
Pittsburgh	84	84	74	84
Seattle	80	76	74	77
Spokane	66	38	49	56
	$1 - \left( \frac{\text{var}(\beta_M^{NY}   \text{data}_M^{NY}, \text{data})}{\text{var}(\beta_M^{NY}   \text{data})} \right)$	$1 - \left( \frac{\text{var}(\beta_H^{NY}   \text{data}_M^{NY}, \text{data})}{\text{var}(\beta_H^{NY}   \text{data})} \right)$	$1 - \left( \frac{\text{var}(\beta_M^{NY}   \text{data}_M^{NY}, \text{data}_M)}{\text{var}(\beta_M^{NY}   \text{data}_M)} \right)$	$1 - \left( \frac{\text{var}(\beta_H^{NY}   \text{data}_H)}{\text{var}(\beta_H^{NY}   \text{data}_H)} \right)$
New York <sup>b</sup> —prior distributions				
Baseline on $\Sigma$	65	15	67	0
Uniform on $\Sigma$	92	45	—	—
Jeffrey on $\Sigma$	94	49	—	—
Uniform on $B_0$	93	32	—	—

<sup>a</sup> Reductions are those obtained by including time-series data (termed *data* in the equations above) for a particular city (*c*) compared with not including the same data. These reductions were calculated under a combined analysis, as well as under a separate analysis that assumes  $\rho = 0$ . —, not applicable.

<sup>b</sup> For New York, time-series data were available for mortality only.

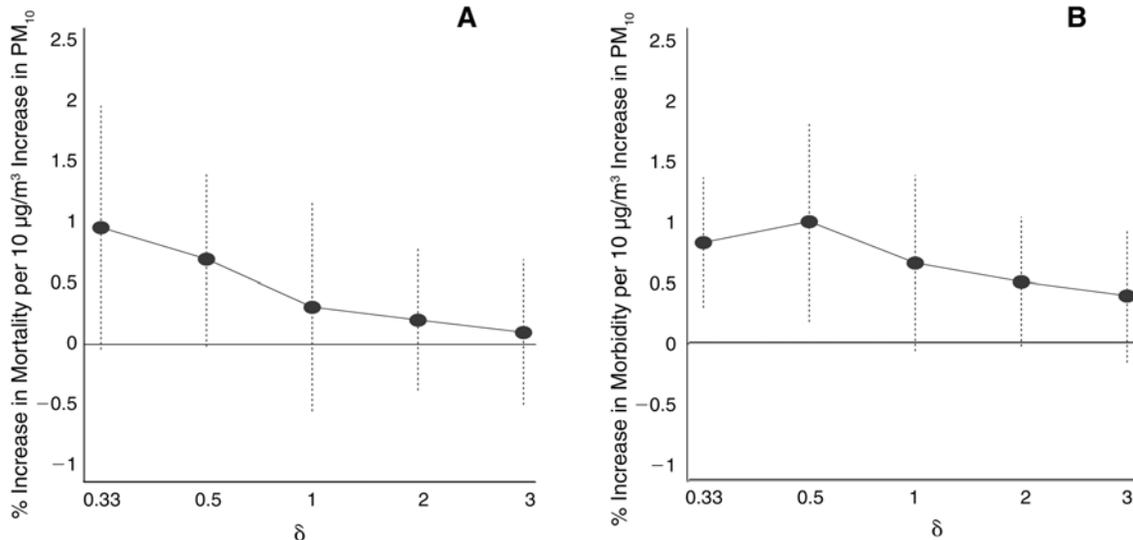


Figure 5. Overall log-relative rates of mortality ( $\alpha_M$ ; A) and hospital admissions ( $\alpha_H$ ; B) plotted in correspondence of five alternative scenarios for adjustment for confounding factors. Calibration parameter ( $\delta$ ) values were used to multiply the number of degrees of freedom in the smooth functions of time, temperature, and barometric pressure. Our baseline model corresponded to  $\delta = 1$ . Vertical dashed lines are 95% posterior intervals.

under the combined analysis than under the separate analysis. This result occurs because in the combined analysis we used data for both mortality and hospital admissions to approximate the marginal posterior distribution of  $\beta_M^c$  (or  $\beta_H^c$ ), whereas under the separate analysis only the mortality (or hospital admission) data were used.

Percentage reductions in posterior variances of these log-relative rates were also greater in the presence of greater heterogeneity across cities as postulated from the prior distribution. This pattern is found because if the prior distribution gives more plausibility to high heterogeneity, then the Bayesian estimate of a city-specific log-relative rate draws more heavily on the data from that city and therefore the reduction in the posterior variances of the two log-relative rates increases. For example, under a uniform prior on  $B_0$  (which leads to larger estimates of  $\sigma_M^2$  and  $\sigma_H^2$ ) the reductions in the posterior variances of  $\beta_M^c$  and  $\beta_H^c$  for New York were 93% and 32%, compared with 65% and 15% under our baseline prior on  $\Sigma$  (for which the estimates of heterogeneity are smaller). Finally, percentage reductions in posterior variances of the log-relative rates of hospital admission and mortality were greater in cities with smaller statistical variances, such as  $\text{var}(\hat{\beta}_H^c | \beta_H^c)$ , of the relative rates. For example, in Detroit, where  $\text{var}(\hat{\beta}_H^c | \beta_H^c) = 0.22$ , we estimated an 87% reduction in the posterior variance of  $\beta_H^c$  under the combined analysis, whereas in Canton where  $\text{var}(\hat{\beta}_H^c | \beta_H^c) = 1.04$ , we estimated a 26% reduction, because of the great statistical uncertainty in  $\hat{\beta}_H^c$ .

We found that the overall log-relative rate of mortality is sensitive to the degree of adjustment for confounding factors: it loses significance when  $\delta = 2$  and  $\delta = 3$  (Figure 5). At the other end, the log-relative rate of hospital admission is robust to the specification of  $\delta$ .

## DISCUSSION AND CONCLUSIONS

Although understanding of underlying mechanisms remains limited, abundant observational evidence indicates that current levels of airborne particulate matter are associated with mortality counts and various indexes of morbidity (Pope and Dockery 1999; US Environmental Protection Agency 2004). Time-series analyses have been carried out to characterize the effect of particulate matter on a variety of health outcomes (including mortality, hospitalization, emergency room visits, and clinic or physician visits). In general, results from separate analyses link particulate matter to increased risk for each of these outcomes. The cities included in each of these different analyses overlap somewhat; for example, by design, some of the same cities were included in morbidity and mortality analyses for NMMAPS and the Air Pollution and Health: A European Approach (APHEA) study (Katsouyanni et al 1997, 2001). However, patterns of effect correlations among the cities for different health outcomes have not yet been examined.

Numerous hypotheses have been proposed to explain the nature of injury processes induced by particulate matter. In general, the same pathogenetic mechanisms

have been considered responsible for effects on mortality or morbidity. Additionally, the same populations have been considered susceptible to the effects of particles: infants, older persons, and persons with chronic cardiac and respiratory diseases. Air pollution has been thought to worsen their clinical condition and thereby increase their risk of hospitalization and, ultimately, death. These biomedical considerations imply that the effects of particulate air pollution on morbidity and mortality should be correlated. The present analyses provided insufficient evidence to test for these hypothesized correlations.

The reason for expecting a correlation between the effects of  $PM_{10}$  on admissions and deaths is not that the same individuals are being affected. We have no evidence to support the hypothesis of a lagged *within-city correlation* between hospital admissions (as the response variable) and mortality. The view that such a lag exists reflects the belief that air pollution effects cause hospitalization first, mortality second. Empirical evidence suggests that this belief is not correct.

Schwartz (1994b) reported that, of all deaths, out-of-hospital sudden deaths and dead-on-arrival deaths increased most on high compared with low air pollution days. These types of deaths represented most of the pollution-related excess deaths. These persons did not first get admitted and subsequently die. This fact was confirmed in a later study, in which Poisson regressions of daily death counts in 10 US cities were fit, stratified by whether deaths were in or out of hospital (Schwartz 2000). The combined regression coefficient for in-hospital deaths was less than one-third of the coefficient for out-of-hospital deaths.

Exposure measurements also suggest that observed deaths are not predominantly among people who had first been admitted to the hospital. Most US hospitals have air filtration systems. Zanobetti and colleagues (2004) measured  $PM_{2.5}$  concentrations in the cardiac rehabilitation unit of the Brigham and Women's Hospital (Boston MA) and found them to be extremely low compared with outdoor  $PM_{2.5}$  concentrations. In contrast,  $PM_{2.5}$  levels within homes were not low and were well correlated with outdoor air. Hence patients in hospitals are unlikely to be breathing polluted air.

Motivated by these general pathogenetic considerations, we developed a hierarchical bivariate time-series model to jointly assess the relations between mortality and hospital admission in 10 US cities. These cities were selected on the basis of the availability of  $PM_{10}$  concentration, hospitalization, and mortality data; they are representative of neither the NMMAPS data nor of the United States. Nonetheless, the cities have varying characteristics and are scattered across the United States. From them, we were able to

gain insights concerning the correlation between log-relative rates of hospital admission and mortality among cities.

Our modeling approach extends Poisson regression analyses of univariate time-series data on air pollution and health to multivariate health outcomes. Within each city, two seemingly unrelated Poisson regression models were fit to estimate log-relative rates of hospital admission and mortality ( $\hat{\beta}_H^c$  and  $\hat{\beta}_M^c$ ). In addition, a novel generalized estimating equation approach was developed to estimate the sample covariance matrix of the relative rates ( $V^c$ ) by using the bivariate time series on hospital admission and mortality  $[y_{IH}^c, y_{IM}^c]$  directly. This bivariate Poisson time-series model was then extended in a hierarchical fashion to combine across cities the vectors of the city-specific estimates of the relative rates of mortality and hospital admission. Although taking into account the correlation between the mortality and hospital admission time series within each city is important, we focused our analysis on (1) making inferences about the parameters in the second stage of the hierarchical models and (2) approximating the marginal posterior distributions of the overall log-relative rates of mortality and hospital admission, their between-city variances, and their correlation across cities.

The combined analysis approach has several useful features.

- By estimating the covariances between the log-relative rates ( $v_{HM}^c$ ), it takes into account correlation between the mortality and hospital admission time series.
- It provides more efficient estimates of the relative rates than would separate analyses, because it uses data for both mortality and hospital admission to approximate the marginal posterior distribution of  $\beta_H^c$  (or  $\beta_M^c$ ), whereas under the separate analysis only the hospital admission (or mortality) data are used.
- It can be used to predict a hospitalization log-relative rate for an additional city by using mortality data; such a prediction might be useful for policy purposes.
- For a city of interest  $c'$ , it quantifies the reduction in the variances of  $\beta_H^{c'}$  and  $\beta_M^{c'}$  that would be obtained by collecting time-series data for  $c'$  compared with predicting  $\beta_H^{c'}$  and  $\beta_M^{c'}$  on the basis of the data from other cities only.

Application of our two-stage bivariate normal-normal model to 10 time series of daily mortality and hospital admission data provided estimates consistent with results of previous studies that evaluated hospital admission and mortality separately. Overall log-relative rates of mortality and hospital admission obtained by combining information across the 10 cities were similar to those reported in the recent NMMAPS revised analyses for 90 and 14 cities

(Dominici et al 2003; Schwartz et al 2003), but they had larger posterior intervals due to the smaller number of cities analyzed here. As expected, the overall log-relative rate of hospital admissions was greater and less heterogeneous than the overall log-relative rate of mortality. Because of the large statistical uncertainty within each city, the correlation coefficient  $\rho$  was estimated poorly, thus providing little information about the overall association between log-relative rates of mortality and hospital admission. We should address this issue by applying our modeling strategy to longer time series for pollution, mortality, and hospital admission for a larger number of cities. Because of the biological basis of a correlation between morbidity and mortality health effects, as well as our initial findings, such additional exploration is needed.

Recent advances in methods for semiparametric regression could also be used to extend our modeling approach. For example, we could have used a Bayes approach via Markov chain Monte Carlo (MCMC) sampling for inference in generalized additive models with city-specific random effects, as previously suggested (Fahrmeir and Tutz 2001). With this approach, we would not have needed to assume normality in the first stage of the hierarchical model. However, to properly estimate the sample covariance between the estimated log-relative rates for mortality and hospital admission, extending generalized additive models with random effects to multivariate outcomes is needed.

---

## ACKNOWLEDGMENTS

Partial funding for Francesca Dominici was provided by a National Institute of Environmental Health Sciences RO1 grant (ES012054-01). Funding was also provided by that Institute's Center in Urban Environmental Health (P30 ES 03819).

---

## REFERENCES

Albert PS. 1999. Longitudinal data analysis (repeated measures) in clinical trials. *Stat Med* 18:1707–1732.

Bascom R, Bromberg PA, Costa DA, Devlin R, Dockery DW, Frampton MW, Lambert W, Samet JM, Speizer FE, Utell M (Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society). 1996. Health effects of outdoor air pollution, Part 1. *Am J Respir Crit Care Med* 153:3–50.

Bascom R, Bromberg PA, Costa DL, Devlin R, Dockery DW, Frampton MW, Lambert W, Samet JM, Speizer FE, Utell M (Committee of the Environmental and Occupational Health

Assembly of the American Thoracic Society). 1996b. Health effects of outdoor air pollution, Part 2. *Am J Respir Crit Care Med* 153:477–498.

Breslow N, Clayton D. 1993. Approximation inference in generalized linear mixed models. *J Am Stat Assoc* 88:9–25.

Carlin BP, Louis, TA. 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, New York NY.

Daniels M. 1999. A prior for the variance in hierarchical models. *Can J Stat* 27:569–580.

Daniels M, Dominici F, Samet JM, Zeger SL. 2000. Estimating particulate matter-mortality dose-response curves and threshold levels: An analysis of daily time-series for the 20 largest US cities. *Am J Epidemiol* 152:397–412.

Daniels MJ, Dominici F, Zeger SL, Samet JM. 2004. The National Morbidity, Mortality, and Air Pollution Study, Part III: Concentration–Response Curves and Thresholds for the 20 Largest US Cities. Research Report 94. Health Effects Institute, Boston MA.

Daniels M, Kass R. 1999. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J Am Stat Assoc* 94:1254–1263.

Dockery D, Pope CA. 1994. Acute respiratory effects of particulate air pollution. *Annu Rev Public Health* 15:107–132.

Dockery D, Pope CA III, Xu X, Spengler J, Ware J, Fay M, Ferris B, Speizer F. 1993. An association between air pollution and mortality in six US cities. *N Engl J Med* 329:1753–1759.

Dominici F, Daniels M, Zeger SL, Samet JM. 2002. Air pollution and mortality: Estimating regional and national dose-response relationships. *J Am Stat Assoc* 97:100–111.

Dominici F, McDermott A, Daniels M, Zeger SL, Samet JM. 2003. Mortality among residents of 90 cities. In: *Revised Analyses of Time-Series Studies of Air Pollution and Health*, pp 9–24. Special Report. Health Effects Institute, Boston MA.

Dominici F, McDermott A, Hastie TJ. 2004a. Improved semiparametric time series models of air pollution and mortality. *J Am Stat Assoc* 99(468):938–948.

Dominici F, Samet JM, Zeger SL. 2000. Combining evidence on air pollution and daily mortality from the twenty largest US cities: A hierarchical modeling strategy. *R Stat Soc Ser A* 163:263–302.

Dominici F, Zanobetti A, Zeger SL, Schwartz J, Samet JM. 2004b. Hierarchical bivariate time series models: A combined

- analysis of the effects of particulate matter on morbidity and mortality. *Biostatistics* 5:341–360.
- DuMouchel W. 1990. *Bayesian Metaanalysis*. Marcel Dekker, New York NY.
- DuMouchel WH, Harris JE. 1983. Bayes methods for combining the results of cancer studies in humans and other species. *J Am Stat Assoc* 78:293–308.
- Environmental Protection Agency (US). 2004. *Epidemiology of human health effects associated with ambient particulate matter*. Chapter 8 in: *Air Quality Criteria for Particulate Matter, Volume II*. EPA/600/P-99/002bF. Office of Research and Development, Research Triangle Park NC.
- Everson P, Morris C. 2000. Inference for multivariate normal hierarchical models. *J R Stat Soc Ser B* 62:399–412.
- Fahrmeir L, Tutz G. 2001. *Multivariate Statistical Modeling Based on Generalized Linear Models*. Springer-Verlag, New York NY.
- Gelman A, Carlin J, Stern H, Rubin D. 1995. *Bayesian Data Analysis*, pp 320–362. Chapman & Hall, London England.
- Gilks WR, Richardson S, Spiegelhalter DJ, eds. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, England.
- Hastie TJ, Tibshirani RJ. 1990. *Generalized Additive Models*. Chapman & Hall, New York NY.
- Hwang J-S, Chan C-C. 2002. Effects of air pollution on daily clinic visits for lower respiratory tract illness. *Am J Epidemiol* 155:1–10.
- Katsouyanni K, Toulomi G, Samoli E, Gryparis A, LeTertre A, Monopoli Y, Rossi G, Zmirou D, Ballester F, Boumghar A, Anderson HR. 2001. Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 project. *Epidemiology* 12:521–531.
- Katsouyanni K, Toulomi G, Spix C, Balducci F, Medina S, Rossi G, Wojtyniak B, Sunyer J, Bacharova L, Schouten J, Ponka A, Anderson HR. 1997. Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. *Br Med J* 314:1658–1663.
- Kelsall J, Samet JM, Zeger SL. 1997. Air pollution and mortality in Philadelphia, 1974–1988. *Am J Epidemiol* 146:750–762.
- Korrick S, Neas L, Dockery D, Gold D. 1998. Effects of ozone and other pollutants on the pulmonary function of adult hikers. *Environ Health Perspect* 106:93–99.
- Li Y, Roth H. 1995. Daily mortality analysis by using different regression models in Philadelphia county, 1973–1990. *Inhal Toxicol* 7:45–58.
- Liang K-Y, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22.
- Lipfert F, Wyzga R. 1993. Air pollution and mortality: Issues and uncertainty. *J Air Waste Manag Assoc* 45:949–966.
- McNeney B, Petkau J. 1994. Overdispersed poisson regression models for studies of air pollution and human health. *Can J Stat* 22:421–440.
- Pope CA. 2000. Invited commentary: Particulate matter–mortality exposure–response relations and threshold. *Am J Epidemiol* 152:407–412.
- Pope CA, Dockery DW. 1999. Epidemiology of particle effects. In: *Air Pollution and Health*, pp 673–705. Academic Press, San Diego CA.
- Rochon J. 1996. Accounting for covariates observed post randomization for discrete and continuous repeated measures data. *J R Stat Soc Ser B* 58:205–219.
- Samet JM, Zeger SL, Berhane K. 1995. The Association of Mortality and Particulate Air Pollution. In: *Particulate Air Pollution and Daily Mortality: Replication and Validation of Selected Studies; the Phase I Report of the Particle Epidemiology Evaluation Project*. Special Report. Health Effects Institute, Cambridge MA.
- Samet JM, Zeger SL, Kelsall JE, Xu J, Kalkstein LS. 1997. *Particulate Air Pollution and Daily Mortality: Analyses of the Effects of Weather and Multiple Air Pollutants*. The Phase I.B Report of the Particle Epidemiology Evaluation Project. Special Report. Health Effects Institute, Cambridge MA.
- Samet JM, Dominici F, Curriero F, Coursac I, Zeger SL. 2000a. Fine particulate air pollution and mortality in 20 US cities: 1987–1994. *N Engl J Med* 343(24):1742–1749.
- Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000b. *The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues*. Research Report 94. Health Effects Institute, Cambridge MA.
- Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, Dockery DW, Schwartz J, Zanobetti A. 2000c. *The National Morbidity, Mortality, and Air Pollution Study, Part II: Morbidity and Mortality from Air Pollution in the United*

States. Research Report 94. Health Effects Institute, Cambridge MA.

Schwartz J. 1994a. Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Can J Stat* 22:471–488.

Schwartz J. 1994b. What are people dying of on high air pollution days? *Environ Res* 64:26–35.

Schwartz J. 1995. Air pollution and daily mortality in Birmingham, Alabama. *Am J Epidemiol* 137:1136–1147.

Schwartz J. 2000. Assessing confounding, effect modification, and thresholds in the association between ambient particles and daily deaths. *Environ Health Perspect* 108:563–568.

Schwartz J, Zanobetti A. 2000. Is there a threshold in the relation between particulate air pollution and daily deaths: A meta-smoothing approach. *Epidemiology* 6:666–672.

Schwartz J, Zanobetti A, Bateson T. 2003. Morbidity and mortality among elderly residents of cities with daily PM measurements. In: *Revised Analyses of Time-Series Studies of Air Pollution and Health*, pp 25–58. Special Report. Health Effects Institute, Boston MA.

Zanobetti A, Canner MJ, Stone PH, Schwartz J, Sher D, Eagan-Bengston E, Gates KA, Hartley LH, Suh H, Gold DR. 2004. Ambient pollution and blood pressure in cardiac rehabilitation patients. *Circulation* 110(15):2184–2189.

Zanobetti A, Schwartz J, Dockery D. 2000a. Airborne particles are a risk factor for hospital admissions for heart and lung disease. *Environ Health Perspect* 108:1071–1077.

Zanobetti A, Wand M, Schwartz J, Ryan L. 2000b. Generalized additive distributed lag models. *Biostatistics* 1:279–292.

Zeger SL, Dominici F, Samet JM. 1999. Harvesting-resistant estimates of pollution effects on mortality. *Epidemiology* 89:171–175.

Zeger SL, Liang K-Y, Albert PS. 1988. Models for longitudinal data: A generalized estimating equation approach. *Biometrics* 44:1049–1060 (correction, *Biometrics* 45:347).

Zeger SL, Liang K-Y. 1992. An overview of methods for the analysis of longitudinal data. *Stat Med* 11:1825–1839.

---

## ABOUT THE AUTHORS

---

**Francesca Dominici** is associate professor of biostatistics at the Johns Hopkins Bloomberg School of Public Health. Dr Dominici received her doctoral degree from the University of Padua. Her research interests include time-series analysis in environmental epidemiology, semiparametric Bayesian models for dose–response data, hierarchical models for combining information and for meta-analysis, and missing data. She is the first recipient of HEI’s Walter A Rosenblith New Investigator Award. Her work on air pollution and mortality was honored by an invitation to read her paper on this topic to the Royal Statistical Society in London.

**Antonella Zanobetti** received her PhD in 1999 in applied statistics from the University of Florence. She is currently a research associate in the Exposure Epidemiology and Risk Program at Harvard School of Public Health. Her field of research includes statistical analysis of environmental epidemiology data with particular interest in adverse effects of air pollution on health. Her expertise is analyzing the short-term effect of air pollution on mortality and morbidity by means of time-series analysis. Dr Zanobetti’s other research interests include mechanisms linking inhalation of ambient particles to acute exacerbation of cardiovascular or respiratory disease; health effects of air pollution and temperature extremes; and socioeconomic influences on health. She is also interested in developing innovative statistical methodologies in environmental epidemiology.

**Scott L Zeger** is professor and chair of the Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health. Dr Zeger earned his doctorate in statistics from Princeton University. With Kung-Yee Liang, he developed the generalized estimating equation approach to regression modeling with correlated data as occurs in time series or longitudinal studies. This work has been recognized with awards from both the American Statistical Association and American Public Health Association. Dr Zeger’s current research focuses on improving the design and analysis of data from prospective studies.

**Joel Schwartz**, professor of environmental epidemiology at Harvard School of Public Health and Harvard Medical School, received his doctorate from Brandeis University in theoretical physics and the John D and Catherine T MacArthur Fellowship award. He has held appointments as a visiting scientist at the University of Basel, Switzerland, and at the University of Wuppertal, Germany, and has also served as a senior scientist at the US Environmental Protection Agency. His research focuses on the effects of antioxidants

on respiratory health, the health effects of air and water pollution, lead, and methodologic questions regarding the modeling of covariates in epidemiologic studies. Over more than a decade, his studies have been instrumental in precipitating renewed interest in airborne particles.

**Jonathan M Samet** is professor and chair of the Department of Epidemiology at the Johns Hopkins Bloomberg School of Public Health. He received his medical degree from the University of Rochester School of Medicine and Dentistry and a master's degree in epidemiology from the Harvard School of Public Health. His research has focused on the effects of environmental and occupational agents. He was chair of the National Research Council's Committee on Research Priorities for Airborne Particulate Matter and has served as chair of the Council's Committee on Health Risk of Exposure to Radon (BEIR VI). He has served on the Clean Air Scientific Advisory Committee (CASAC) of the Environmental Protection Agency. In October 2000, Dr Samet was installed as the first Jacob I and Irene B Fabrikant Professor of Health, Risk, and Society at the Johns Hopkins University.

---

#### OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

---

Dominici F, Zanobetti A, Zeger SL, Schwartz J, Samet JM. 2004. Hierarchical bivariate time series models: A combined analysis of the effects of particulate matter on morbidity and mortality. *Biostatistics* 5(3):341–360.

---

#### ABBREVIATIONS AND OTHER TERMS

---

<i>df</i>	degrees of freedom
NMMAPS	National Morbidity, Mortality, and Air Pollution Study (US)
NY	New York [city]
PM	particulate matter
PM <sub>10</sub>	particulate matter < 10 µm in aerodynamic diameter



## INTRODUCTION

Over more than a decade, time-series epidemiologic studies have been conducted in many cities to evaluate the association between daily changes in concentrations of particulate matter in the air and daily counts of morbidity and mortality. The National Morbidity, Mortality, and Air Pollution Study (NMMAPS\*), which HEI has funded since 1996<sup>†</sup>, was designed to address concerns about bias in the selection of cities in air pollution studies, to allow effects of PM to be estimated more precisely, and to explore heterogeneity of effects. To accomplish these objectives, NMMAPS investigators at Johns Hopkins University and Harvard University selected multiple cities according only to population size and the availability of monitoring data for particulate matter less than 10  $\mu\text{m}$  in aerodynamic diameter (PM<sub>10</sub>).

Epidemiologic observational studies as a whole have largely supported an association between increases in pollution and mortality (Health Effects Institute 2003). Although a causal association cannot be inferred from the mere presence of an association based on observational data, it can be argued on this basis. The Bradford Hill criteria, used to interpret evidence from epidemiologic and other types of studies, were motivated by the perceived need to itemize the considerations on which a causal argument could be based (Hill 1965). Hill did not intend these to be used as sine qua non criteria for inference, and their application in inferring causation has exceptions (Rothman 1986). Nevertheless, the Hill criteria have been applied to air pollution–mortality associations to make a case for causation (Ostro 1993).

One of the nine criteria proposed by Hill was coherence. Coherence indicates that a causal basis for an association does not conflict with what is known of the biology or natural history of the disease in question. In the context of air

pollution, coherence has been used in a more particular sense than that originally proposed by Hill: Findings have been termed coherent if similar findings are present for a group of outcomes that should be related (Bates 1992). The pyramid of adverse health effects has been used as a graphical aid to depict relations among outcomes by degree of adversity (American Thoracic Society 1985). The most adverse outcome, mortality, occupies the top of the pyramid, with less adverse outcomes (from greater to lesser adversity) farther down. An argument for causation based in part on coherence in this relatively narrow sense would require that associations of pollution with more adverse health outcomes be mirrored in associations with less adverse health outcomes.

Using this sense of coherence as part of a broader discussion of causation would require specifically that associations between short-term increases in air pollution concentrations and mortality also be seen for hospitalizations, an indicator of morbidity. This is the sense of coherence that motivated HEI to support the work in NMMAPS IV. Furthermore, associations should also be seen for less adverse outcomes (such as emergency room and doctor's office visits). That is, findings on mortality should merely represent the tip of the iceberg. On the basis of published results, Bates has argued that PM is associated with mortality, hospitalizations, symptoms, and medication use (Bates 1992). Coherence may therefore exist for air pollution and health effects, which would strengthen the case for causation.

An argument for this definition of coherence does not require any formal or quantitative assessment, such as a requirement that cities with strong PM–mortality associations also have strong associations between PM and hospitalizations and other outcomes (or even that associations of any magnitude for a range of outcomes need to be present within the same city). To claim coherence for associations of PM with mortality, it is sufficient that associations with less adverse outcomes also be reported in the literature.

Coherence can also be considered within a context broader than just observational studies. If the PM association with mortality is indeed causal, coherence is expected not only among observational findings but also between observational and nonobservational findings. The latter sense of coherence is most in keeping with Hill's use of the term. Therefore, findings from toxicologic and human experimental studies should also be used to judge whether coherence exists.

Findings from NMMAPS support an association between short-term increases in the concentration of PM less than

\* A list of abbreviations and other terms appears at the end of the Investigators' Report.

<sup>†</sup> The "National Morbidity, Mortality, and Air Pollution Study," Dr Jonathan M Samet, principal investigator, began in December 1996 and has cost about \$976,000 for all reports (Parts I–IV). The draft Part IV Investigators' Report was received for review in July 2003. A revised report, received in June 2004, was accepted for publication in December 2004. During the review process, the NMMAPS Special Panel of the HEI Health Review Committee and the investigators had the opportunity to exchange comments and to clarify issues in both the Investigators' Report and in the Review Committee's Commentary.

This document has not been reviewed by public or private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views of these parties, and no endorsements by them should be inferred.

10  $\mu\text{m}$  in aerodynamic diameter ( $\text{PM}_{10}$ ) and mortality among all residents in the 90 largest cities in the United States where PM was monitored at least every 6 days (Samet et al 2000b; Dominici et al 2003). The NMMAPS investigators reported similar findings for hospitalizations for cardiovascular and respiratory diseases and mortality among residents 65 years of age and older in a smaller group of cities with daily  $\text{PM}_{10}$  monitoring (Samet et al 2000b; Schwartz et al 2003). They realized that the latter findings, which involve both mortality and hospitalizations, could be used to conduct a parallel assessment of how an underlying process affects two outcome measures.

Rather than simply noting that associations were present for both mortality and hospitalizations, however, the NMMAPS investigators elected to use a formal quantitative approach that entailed evaluating whether cities with large  $\text{PM}_{10}$  effect estimates for mortality also have large  $\text{PM}_{10}$  effect estimates for hospitalizations, and vice versa: that is, whether the size of effect estimates for mortality and hospitalizations within cities is related. By conducting such an assessment they evaluated coherence within the observational setting, which was one of HEI's original objectives in funding NMMAPS. This quantitative evaluation of coherence is more stringent than that outlined by Bates (1992); not only does it require that effects for outcomes expected to show similar associations with PM be present within the same cities, but also that the magnitudes of these associations are related.

To conduct this evaluation, the NMMAPS investigators limited their analysis to the small subset of cities that was common to the mortality and hospitalization analyses (Samet et al 2000b; Schwartz et al 2003). They also used a standardized analytic approach, even though originally mortality and hospitalization data in these cities were analyzed using somewhat different approaches (Samet et al 2000b; Schwartz et al 2003). Any lack of correlation could not, then, be attributed to differences in analytic approach.

Because the methods described in the Investigators' Report are complex, in this Commentary we describe them in somewhat technical terms and also, in the sidebar, in terms that are more accessible to the general reader.

---

## STUDY OBJECTIVES

---

Several questions could be addressed by the combined analysis of air pollution effects on mortality and on hospital admissions. This study pursued one: Is the underlying

true effect per unit  $\text{PM}_{10}$  on mortality (the mortality slope) of the same magnitude as the effect per unit  $\text{PM}_{10}$  on hospitalizations (the hospitalization slope) in a given city? In other words, are the true mortality slopes and the true hospitalization slopes associated? (The true slope is the slope that would be found in the absence of sampling error, for example, if each city's time series were extremely long.)

To better understand the study and its methods, we describe two other questions targeted by combined analysis that were not addressed in this study but that are similar to the question that was addressed. One of these questions is: To what extent are the individuals who are hospitalized as a result of exposure to  $\text{PM}_{10}$  the same as those who die later as a result of such exposure? This objective could only be studied if hospitalizations were individually matched to deaths—and those data are not available from the NMMAPS database. The other is: To what extent does daily variation in hospitalizations match daily variation in mortality? In other words, how correlated are the two time series within each city? This question could be further refined to consider correlations before and after allowing for the effects of measured determinants (eg, season, weather, pollution). The NMMAPS data could be used to address this question. Although it was not the focus of this study, estimating certain of these within-city correlations was necessary in order to discount their effect when investigating the between-city correlations of mortality and hospitalization slopes.

---

## ANALYTIC METHODS

---

The Investigators' Report presents new statistical methods that were designed to jointly model the effect of air pollution on both mortality and hospitalizations from several locations (in this study, cities) to combine information across locations in a meaningful way.

The statistical modeling was conducted in two stages. In the first stage, the following steps were performed. For each city, two models were applied—one for mortality and one for hospitalizations—by using well-established methods that take autocorrelation, long-term time trends, seasonality, and confounding into account by applying smoothing techniques. These methods are based on Poisson regression, which allows for overdispersion. (They have been developed and described by the authors in a series of publications, including previous NMMAPS reports [Samet et al 2000a,b].) The result of this analysis

was to estimate for each city the association of PM<sub>10</sub> with mortality and with hospitalizations, separately. These estimated associations were expressed as log-relative rates.

The correlation between the estimated association of PM<sub>10</sub> concentration with mortality and the estimated association of PM<sub>10</sub> concentration with hospital admissions was then estimated by using a new method based on generalized estimating equations. The numerical calculation of the estimated correlation uses a weighted average of the cross-correlation between the mortality and hospitalization time series in each city.

In the second stage, information was combined across cities. The result was a combined estimate of the association of PM<sub>10</sub> and mortality for all 10 cities and a combined estimate of the association of PM<sub>10</sub> and hospital admissions for all 10 cities. The method chosen for this combination is based on a Bayesian argument and an assumption of normality. The main assumption is that the pair of log-relative rates (for mortality and for hospital admissions) follows a normal distribution. The mean vector of the normal distribution represents the average of the PM<sub>10</sub> association with mortality and the PM<sub>10</sub> association with hospital admissions for all cities; the variances of the normal distribution

represent the precision with which these associations can be estimated; and the assumption of a nonzero covariance allows for a common correlation between the mortality and hospital admissions log-relative rates. Thus, this second stage models the variation and covariation between the true relative rates associated with mortality and hospital admissions across cities.

The method used to estimate the parameters of this normal distribution is a standard simulation technique, widely used in Bayesian models of this type, called *Markov chain Monte Carlo sampling*. This sampling method provides estimated *posterior* distributions for the unknown parameters. Moreover, it permits more refined estimation of the PM<sub>10</sub> associations in each city by borrowing strength from the other cities (which figured prominently in NMMAPS II [Samet et al 2000b; Dominici et al 2003]).

These new statistical methods are innovative because they appropriately model two separate outcomes (ie, mortality and hospitalization counts) from several locations and estimate log-relative rates, taking into account the multiple sources of dependence within and between locations.

The Bayesian hierarchical procedure requires that prior distributions be assigned for each parameter. To avoid undue

## ANALYTIC METHODS FOR THE NONSTATISTICIAN

### Why New Statistical Methods Were Necessary

At first glance the task—investigating the association between 10 mortality slopes and 10 hospitalization slopes—may seem simple. Why not just calculate a correlation or regression coefficient between the two sets of slopes? Unfortunately, this solution ignores the statistical, or sampling, error in the estimated slopes. For example, even if the underlying true slopes were perfectly correlated across cities, the errors in estimating the slopes will add noise that will render the correlation between estimated slopes imperfect.

Further, the sampling errors in the mortality and hospitalization slopes in a city are not independent. A sampling error that artificially elevates a mortality slope might also give rise to a higher hospitalization slope. Unless this correlation between sampling errors is taken into account in the data analysis, it could cause at least part of a positive association between the mortality and hospitalization slopes across the 10 cities.

Thus, estimation of the underlying association between mortality and hospitalization slopes is not straightforward. The technicalities of separating chance from an underlying association have required innovation in statistical method, resulting in the considerably complex approach used in NMMAPS IV.

### General Strategy

The statistical analysis essentially has three steps:

1. Estimate PM<sub>10</sub>-mortality and PM<sub>10</sub>-hospitalization slopes for each city from two seemingly unrelated time-series regressions.

2. Estimate, for each city, the dependence between the mortality and hospitalization slopes expected by **sampling error alone**.
  3. Estimate the underlying association across cities between true mortality and hospitalization slopes, discounting the sampling error in each.
- All three steps use quite complex statistical methods, but only the second uses new methods.

**Step 1: Seemingly Unrelated Regressions** The approach adopted by the investigators starts by estimating for each city a pollution-mortality and a pollution-hospitalization slope

$$\left( \hat{\beta}_M^c \text{ and } \hat{\beta}_H^c, \text{ for cities } c = 1 \dots 10 \right)$$

in conventional time-series regression analyses. These regression models are themselves complex, but the complexities are not new. The investigators have drawn on their previous work to make specific model choices in this step (Investigators' Report Table 2).

Because the two regressions (for mortality and hospitalizations) were fitted separately but the investigators were aware of the hidden dependence between the slope estimates, they refer to the regressions as *seemingly unrelated*. The regressions appear unrelated but, in fact, dependence in the data relates them.

*Continued on next page*

reliance on strong prior assumptions, the investigators chose vague priors for the means, the variance components, and the correlation coefficient. Specifically, they used a normal prior for the means but one with zero means and large variances. In Bayesian parlance, using a large variance for a normal prior distribution is equivalent to imposing a minimal prior belief on the distribution of the parameter of interest. For the variance components, they used half-normal distributions on  $(0, \infty)$  with zero means and large variances. For the correlation coefficient, they used a uniform distribution on the interval  $(-1, 1)$ .

A Bayesian hierarchical model postulates a common mean across cities for the true log-relative risk of pollution. Under this assumption, the posterior distribution of each city-specific log-relative risk is a weighted average of that city's estimate and the prior mean. Because the prior mean is common to all cities, this assumption has the effect of shrinking the city-specific estimates toward a common value. The investigators added a new feature to this hierarchical model: an assumption that the log-relative risks for mortality and hospitalizations are correlated. This assumed correlation is involved in both stages of the model. In the first

### ANALYTIC METHODS FOR THE NONSTATISTICIAN (continued)

**Step 2: Estimated Sampling Dependence Between Estimated Mortality and Hospitalization Slopes** The investigators then needed to estimate the sampling error in each city's pollution–mortality slope and pollution–hospitalization slope. Because the two estimates are not independent, this task required estimating not only the standard errors of each slope

$$\left(\sqrt{v_M^c} \cdot \sqrt{v_H^c}\right),$$

which can be done routinely by using standard theory and software, but also the degree of dependence between the two slopes (based on correlation or covariance due to sampling error,  $v_{MH}^c$ ). This correlation cannot be estimated by using standard software or even standard theory. Estimating this correlation was therefore the principal methodologic innovation and the primary subject of this study.

In order to estimate this sampling dependence of estimated mortality and hospitalization slopes, the investigators had to estimate, in each city, the degree of dependence between the daily mortality and hospitalization counts after allowing for known determinants modeled in the time-series regressions. Specifically, correlations were calculated between the daily residuals in the mortality and hospitalization regressions (shown in Table 3 in the Investigators' Report). These correlations were then used, following a generalized estimating equation approach, to infer the degree of dependence between the pollution–mortality and pollution–hospitalization slopes. Dependence was taken into account in the data analysis not only between the mortality and hospitalization counts on the same day (lag 0), but also with delays (lags) of up to 14 days. The investigators assumed no dependence beyond 14 days. Aside from the choice of this maximum lag of residual dependence (14 days), the methods also required specification of a smoothing function, because without it the dependence had too many parameters to be estimable.

### Step 3: Discounting Sampling Error When Estimating Association Between True Mortality and Hospitalization Slopes Across Cities

The complete sampling error of each city's mortality slope and hospitalization slope (two standard errors and a covariance) was described in a sampling variance–covariance matrix,  $V^c$ :

$$V^c = \begin{bmatrix} v_M^c & v_{MH}^c \\ v_{MH}^c & v_H^c \end{bmatrix}$$

Once the  $V^c$  matrices have been estimated, these sampling errors can be discounted from the association between the 10 estimated mortality and hospitalization slopes to estimate the underlying true association across the 10 cities between true mortality and true hospitalization slopes. The investigators were thus able to estimate the true correlation between the two slopes across cities, allowing for artifacts created by noise.

Technically, this final step required a hierarchical model to separate the sampling variability and correlation of the true slopes ( $\beta_M^c$  and  $\beta_H^c$ ) across cities from the variability and correlation of the estimated slopes ( $\hat{\beta}_M^c$  and  $\hat{\beta}_H^c$ ) in each city. The variation and correlation between true slopes—described in the variance–covariance matrix  $\Sigma$  in expression (3) in the Investigators' Report—together with the slope means,  $\mu_M$  and  $\mu_H$ —are the features of interest at the top of the hierarchy. The sampling variability and correlation of observed slopes,  $V^c$  in expression (2), are nuisance factors at the bottom of the hierarchy:

$$\begin{bmatrix} \hat{\beta}_M^c \\ \hat{\beta}_H^c \end{bmatrix} \sim N_2 \left( \begin{bmatrix} \beta_M^c \\ \beta_H^c \end{bmatrix}, V^c \right) \text{ where } V^c = \begin{bmatrix} v_M^c & v_{MH}^c \\ v_{MH}^c & v_H^c \end{bmatrix} \quad (2)$$

Because estimates for the  $V^c$  are now available, they can be discounted when estimating  $\Sigma$ ,  $\mu_M$ , and  $\mu_H$ . The investigators chose to do so by using a Bayesian approach. This application of hierarchical models is not new; it was used in earlier analyses of NMMAPS data (Samet et al 2000b; Dominici et al 2003). Meta-analysis of two-pollutant models, in particular, uses an essentially identical model, with slopes for two pollutants replacing the slopes for mortality and hospitalization in this study (Samet et al 2000b).

### Using Estimated Association Between True Mortality and True Hospitalization Slopes to Refine Slope Estimates in One City

The investigators might have stopped after estimating the correlation between true slopes, which is of intrinsic interest in arguing for coherence of the data, for example. They went on, however, to illustrate the use of this correlation and other parameters estimated from the hierarchical model to refine estimates of pollution–mortality and pollution–hospitalization slopes for specific cities.

This refining procedure is an extension of calculating Bayes slope estimates as a shrunk compromise between an observed city slope and the overall mean, which is a standard practice (Samet et al 2000; Dominici et al 2003). The extension in this study was intended to exploit the correlation between the true mortality and hospitalization slopes, in particular to estimate a specific hospitalization slope. This method provided a refined shrunk Bayes estimate of the hospitalization slope in the sense that the slope was not only a compromise between the city-specific and the overall mean hospitalization slopes across the 10 cities but it also drew on the city-specific and overall mean mortality slopes. The method was applied to estimate a hospitalization slope in New York, a city for which hospitalization data are not available. Had they not used the correlation they calculated between the mortality and hospitalization slopes, the investigators would have had to use the overall mean hospitalization slope over the 10 cities.

stage, the estimates of the log-relative rates for mortality and hospitalizations (ie,  $\hat{\beta}_M^c$  and  $\hat{\beta}_H^c$ ) within each city  $c$  are assumed to be correlated. Then, this estimated correlation is denoted by

$$\frac{v_{MH}^c}{\sqrt{v_M^c v_H^c}},$$

where  $v_{MH}^c$ ,  $v_M^c$  and  $v_H^c$  denote the covariance between the mortality and hospitalization time series, the variance of the mortality series, and the variance of the hospitalization series, respectively. In the second stage of the model, the true log-relative risks across cities are assumed to have a common correlation, denoted as  $\rho$ .

---

## RESULTS

---

There was little or no correlation between the mortality and hospitalization residual time series within each city. The method of generalized estimating equations was used to estimate the correlation between the estimated association of PM<sub>10</sub> and mortality and the estimated association of PM<sub>10</sub> and hospital admissions:  $\hat{\beta}_M^c$  and  $\hat{\beta}_H^c$ , respectively. The estimated correlation between them is reported for each city (Investigators' Report Table 3, column 5). The sampling correlations are small, ranging from  $-0.05$  to  $0.34$  (column 4). These low correlations may have been due to the true relative risks having little correlation or to errors in estimation of the true risks. Little information was available about the estimated marginal posterior distribution of the correlation coefficient between the true log-relative risks across cities, denoted as  $\rho$ , regardless of the prior assumption (Investigators' Report Figure 3). Table 4 in the Investigators' Report presents maximum likelihood estimates and Bayesian estimates of the PM<sub>10</sub> effect on mortality and on hospitalizations for cardiovascular diseases for each city. The city-specific estimates shrank toward a common value owing to the assumption of a common prior mean in the Bayesian hierarchical model (Figure 1 and Table 4). The common prior assumption leads to narrower confidence intervals than when this assumption is not made (Table 4). The narrower intervals result from the more precise estimation of the city-specific log-relative risk, which is made possible by the hierarchical model combining information from all cities.

Bayesian estimates for each city that assumed no correlation between the mortality and hospitalization PM<sub>10</sub> effect estimates were similar to Bayesian estimates from a joint (bivariate) analysis in which these correlations were allowed to influence the estimates. The similarity of the

results reflects the fact that the mortality and hospitalization PM<sub>10</sub> effect estimates were only weakly correlated. A statistical test of whether PM<sub>10</sub> effect estimates differed across cities detected between-city heterogeneity of PM<sub>10</sub> hospitalization effect estimates but none for mortality. The investigators noted that conclusions regarding heterogeneity of effect were likely sensitive to the choice of prior assumptions about the distribution of the covariance matrix used for the Bayesian estimates. The effect estimates themselves were insensitive to the choice of prior, but their SDs, which were similar for mortality and hospitalizations, were sensitive to it (Investigators' Report Figure 2).

The estimated between-city correlation between effect estimates for mortality and hospitalization was low (0.20) but was estimated with a large degree of uncertainty (95% posterior interval  $-0.89, 0.98$ ; Investigators' Report Figure 3). This estimated correlation was insensitive to the choice of prior.

The investigators also used their model to attempt to tackle another problem: estimating a hospitalization effect for a city for which only mortality data are readily available. They used New York as an example. First, they found greater precision in estimating the PM<sub>10</sub> effect on mortality due to cardiovascular diseases when using mortality data from New York as well as the other cities than when merely using data from the other cities (Investigators' Report Figure 4)—not a surprising finding. They also found that using the New York mortality data did not much improve the precision of the estimated PM<sub>10</sub> effect on New York hospitalizations due to cardiovascular diseases compared with merely using the mortality and hospitalization data from the other cities (Figure 4). This result is due to the low correlation between the estimated between-city mortality and hospitalization PM<sub>10</sub> effects.

Investigators' Report Table 5 shows the gains in precision (expressed as percent reductions in posterior variance) for each city's effect estimates due to including that city's data. These gains are reported for analyses that either ignored the between-city correlation between the mortality and hospitalization effects (separate analysis) or incorporated them (combined analysis). As expected, making use of a city's own data improved precision. However, whereas precision of the mortality estimates was always slightly improved by incorporating the between-city correlation (Table 5, column 2 vs column 4), precision of the hospitalization estimates did not always improve (column 3 vs column 5). The degree of improvement in precision also depended on the degree of between-city heterogeneity of effect that was allowed by different priors (see data for New York in Table 5): a uniform prior, for example, gave more plausibility to high heterogeneity and hence led to greater precision. Finally, the improvements in precision for

each city were greater when the certainty with which city effects were estimated was greater. For example, improvements in precision were greater for Detroit than for Canton because estimates of effect for Detroit were more certain.

The final sensitivity analysis was used to determine whether the choice of degrees of freedom for the time and meteorologic smooth functions in the time-series analyses influenced the results. This analysis was prompted by concerns highlighted in the revised analyses of time-series studies, which were performed because programming for generalized additive model (GAM) software was not always appropriate for analysis of these data (Health Effects Institute 2003). The investigators found that the  $PM_{10}$  mortality effect estimate was sensitive to the degree of smoothing: it increased with decreasing degrees of freedom in the smoothing parameters and decreased with increasing degrees of freedom. The  $PM_{10}$  hospitalization effect estimate, however, was relatively insensitive to degree of smoothing (Investigators' Report Figure 5).

---

## CRITIQUE

---

### METHODS

The main contribution of the Investigators' Report is methodologic. To our knowledge, no method for estimating the correlation between effects of a covariate on two parallel time series of counts over multiple locations has previously been proposed. The methods proposed in the Investigators' Report form the first serious attempt at modeling multiple outcomes on time-series data of counts from many cities. Although other methods have been proposed in the past to enable modeling of multiple longitudinal outcomes, they have mainly dealt with multiple outcomes from a single longitudinal data set (eg, Fitzmaurice and Laird 1995; Bandeen-Roche et al 1997; Regan and Catalano 1999). The methods described in this Investigators' Report make possible, for the first time, joint inference of the effect of air pollution ( $PM_{10}$ ) on two outcomes (mortality and hospitalization) from time-series data of counts from several cities. Specifically, the methods involve combining data within and across cities in a way that accounts for interdependence over time and space. The ability to conduct joint inference on these two separate, but related, outcomes aids quantification of the extent of coherence of the data.

The new generalized estimating equation method used to estimate the sampling correlations for each city is outlined in detail in the published article (Dominici et al 2004) based on the Investigators' Report in this volume

and is also described in the Commentary sidebar (see the end of *Step 2*). The method is based on a weighted average of the observed correlations in the residuals of the fitted model. These correlations were weak (Investigators' Report Table 3, columns 6–8), which may be why the estimated sampling correlations (column 4) are small. The statistical properties of the estimating method have not been studied, however, and estimating these sampling correlations with any precision may be difficult.

The sampling correlation estimates are themselves input into the Bayes hierarchical model during the second stage of the analysis, which requires specification of the prior distributions on parameters. The impact of this prior information on the posterior density for  $\rho$  is unclear (Investigators' Report Figure 3). All priors used gave broadly consistent results, but the little information available about the correlation may come from the prior, not the data. Further study of the methods, possibly in simpler settings, would be valuable.

The usefulness of the proposed methods in analyzing the NMMAPS data set is somewhat tempered by the great imprecision in the estimated correlation between the mortality and hospitalization time-series data. Nevertheless, correlations of interest may be able to be estimated more precisely for other data sets.

### RESULTS

The primary aim of the study was to investigate the association between true  $PM_{10}$ -mortality and true  $PM_{10}$ -hospitalization slopes (see sidebar), which the investigators did by estimating the correlation between them across 10 cities. The most critical limitation of the study—apparent in the very wide posterior interval for this correlation ( $-0.89, 0.98$ )—is the uncertainty in this estimate. This interval is barely narrower than the entire possible interval for the correlation coefficient ( $-1, 1$ ). Thus, this study gives little more information about the association between mortality slopes and hospitalization slopes than was already known. The study lacked the power to detect even substantial associations.

The study lacked power and precision in estimating the association of interest for two reasons. First, even without the complications that this study addressed, estimating a correlation from just 10 data points (cities) will be precise only if the underlying correlation is strong. Second, the precision in estimating the association between the underlying true mortality and hospitalization slopes depends on the degree of variation in these slopes across cities. When the study was planned, available evidence suggested that this variation was substantial. However, revised analyses (conducted after programming shortcomings in generalized

additive model software were identified) estimate the variation of underlying slopes to be lower. In this study, point estimates of this variation were quite high relative to their means (SDs for mortality and hospitalization slopes were 0.42 and 0.31, respectively; the corresponding means were 0.26 and 0.71). However, these point estimates were imprecise and the variation in hospitalization slopes could easily be explained by chance ( $P > 0.05$ ). Thus, with only 10 cities in the study and only moderate evidence for variation in the true slopes across the cities, the correlation between true slopes was estimated imprecisely.

Although the correlation between the true mortality and hospitalization slopes is itself of interest, the Investigators' Report demonstrates how it could be used to predict the distribution of the log-relative risk for hospital admissions by using only mortality time-series data. In view of the low correlations among the data, this use is mainly of methodologic interest. The correlation is used to predict the log-relative risk of hospitalizations for a city (eg, New York) as described in Results in the Investigators' Report. The authors also predicted the log-relative rate for mortality for New York with and without including the New York data (Investigators' Report Figure 4A). Comparison of these two predictions reflects the reduction in variance due to postulating a common mean across cities, which is probably of little practical interest. The similarity of the two predicted distributions for hospitalizations (Figure 4B) reflects the weak correlation between mortality and hospital admission series. Estimates of hospitalizations in New York were largely determined from the estimates in the 10 other cities, which had hospital admissions data, and the model assumed that these are all samples from the same normal distribution.

Reductions in estimated variances of city-specific estimates were greater when calculated with that city's information (combined analysis) than without it. This result may reflect both the weakly correlated time series and the increased precision obtained by assuming a common distribution across cities (ie, the model assumptions; Investigators' Report Table 5). Disentangling these effects is difficult because they are confounded.

Although in hindsight the study did not deliver the information that its authors primarily sought, some of the results are of interest. Specifically, the Panel was intrigued by the low correlations for each city between daily mortality and hospitalization counts (Investigators' Report Table 3). Although given only for certain lags, they suggest that the two series are almost totally independent, at least in these cities. Further investigation of the variation between daily mortality and hospitalization counts in these and other cities might elucidate whether it is due to  $PM_{10}$  or to other factors (such as discordance in variations between hospital admissions and mortality).

## NMMAPS IV AND COHERENCE

The substantive question in NMMAPS IV was whether the observational findings provided evidence that the  $PM_{10}$  effects on the evaluated outcomes were related in each city and across cities. Answering this question addresses coherence, which was one of the original objectives of NMMAPS. The study showed no evidence of such correlation, which does not support an argument for causality based on the considerations of coherence. Can we take these findings, and this reasoning, at face value?

A critical question about so-called negative studies (those in which the null hypothesis is not rejected) is whether statistical power was adequate to confidently reject the null hypothesis if it were in fact false. The correlation between the underlying true mortality and hospitalization slopes was very imprecise in this study. Therefore, although mortality and hospitalization  $PM_{10}$  effect estimates may not correlate across cities, we have little confidence in that conclusion.

On the other hand, few NMMAPS IV results support even the presence of such a correlation. The estimated correlation coefficient might have been larger if data from many more cities had been included. But because the estimated correlation for these 10 cities was so small, it would have to have been unusually low in order for a much larger coefficient to be estimated. No evidence suggests that the 10 cities were atypical such that they would yield an unusually low correlation.

Whether the NMMAPS IV findings are to be taken at face value also depends on whether such a quantitative assessment of coherence is appropriate for evaluating coherence. First, this strict approach confines an evaluation of coherence to the observational setting rather than involving the broader context (the total body of evidence on  $PM_{10}$  pollution effects). Second, coherence in the observational setting does appear to require that two outcome  $PM_{10}$  effect estimates across cities (in this case, mortality and hospitalization estimates) be correlated.

City characteristics (eg,  $PM_{10}$  characteristics, population characteristics, or characteristics of the  $PM_{10}$  monitoring network in a city) that determine the size of  $PM_{10}$  effect estimates probably do not differ for mortality and hospitalizations related to cardiovascular disorders. They might differ, however, if the mortality effect were driven by the triggering of lethal cardiac arrhythmias, for example, if the hospitalization effect were driven by other conditions, and if these effects were caused by different  $PM_{10}$  characteristics that varied across cities. City characteristics that determine the size of effect estimates are more likely to differ for other pairs of outcomes, such as mortality and visits to a physician due to asthma.

---

**CONCLUSIONS**

---

1. The main contribution of NMMAPS IV is methodologic.
2. The Investigators' Report describes, for the first time, a method for estimating correlation between effects of a covariate on two parallel time series of counts over multiple cities.
3. The mortality and hospitalization estimates of effect were only weakly correlated, providing no support for coherence when assessed in this way. Because of the imprecise estimate of the correlation (reflected in the very wide posterior interval), however, the question of coherence in the context of observational studies could not be adequately assessed. The broader question of coherence within the complete body of work on PM<sub>10</sub> (including toxicologic and other experimental findings) is not addressed by these methods.
4. Although the methods could be used to predict hospitalization effect estimates from mortality effect estimates, their utility remains to be demonstrated, given the low correlation between the mortality and hospitalization effect estimates in this study.
5. The finding of low correlations between daily counts of deaths and daily counts of hospitalizations is noteworthy. It suggests that most mechanisms causing fluctuations in these counts over time—whether due to PM<sub>10</sub> or other factors—differ for the two outcomes.

---

**ACKNOWLEDGMENTS**

---

The Health Review Committee thanks the NMMAPS Special Panel for their help in evaluating the scientific merit of the Investigators' Report. The Special Panel members are Sverre Vedal (chair), University of Washington; Ben Armstrong, London School of Hygiene and Tropical Medicine; Kiros Berhane, University of Southern California; Edo D Pellizzari, Research Triangle Institute; Nancy Reid, University of Toronto; and Daniel C Tosteson, Harvard Medical School.

The Committee is also grateful to Aaron J Cohen for his oversight of the study, to Cristina I Cann for her assistance in preparing its Commentary, to Jenny Lamont for her science editing of the Research Report and this Commentary, and to Melissa R Harke, Frederic R Howe, Kasey L Oliver, and Ruth E Shaw for their roles in publishing this Research Report.

---

**REFERENCES**

---

- Bandeen-Roche K, Miglioretti DL, Zeger SL, Rathouz PJ. 1997. Latent variable regression for multiple discrete outcomes. *J Am Stat Assoc* 92:1375–1386.
- Bates DV. 1992. Health indices of the adverse effects of air pollution: The question of coherence. *Environ Res* 59:336–349.
- Dominici F, McDermott A, Daniels M, Zeger SL, Samet SM. 2003. Mortality among residents of 90 cities. In: *Revised Analyses of Time-Series Studies of Air Pollution and Health, Special Report*, pp 9–24. Health Effects Institute, Boston MA.
- Dominici F, Zanobetti A, Zeger SL, Schwartz J, Samet JM. 2004. Hierarchical bivariate time series models: A combined analysis of the effects of particulate matter on morbidity and mortality. *Biostatistics* 5:341–360.
- American Thoracic Society (committee: Andrews C, Buist S, Ferris BG [chair], Hackney J, Rom W, Samet J, Schenker M, Shy C, Strieder D). 1985. Guidelines as to what constitutes an adverse respiratory health effect, with special reference to epidemiologic studies of air pollution. *Am Rev Respir Dis* 131:666–668.
- Fitzmaurice GM, Laird NM. 1995. Regression models for a bivariate discrete and continuous outcome with clustering. *J Am Stat Assoc* 90:845–852.
- Health Effects Institute. 2003. *Revised Analyses of Time-Series Studies of Air Pollution and Health, Special Report*. Health Effects Institute, Boston MA.
- Hill AB. 1965. The environment and disease: Association or causation? *Proc R Soc Med* 58:295–300.
- Ostro B. 1993. The association of air pollution and mortality: Examining the case for inference. *Arch Environ Health* 5:336–342.
- Regan MM, Catalano PJ. 1999. Likelihood models for clustered binary and continuous outcomes: Application to developmental toxicology. *Biometrics* 55:760–768.
- Rothman KJ. 1986. *Modern Epidemiology*, pp 16–20. Little, Brown & Co, Boston MA.
- Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000a. The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues. *Research Report* 94. Health Effects Institute, Cambridge MA.

Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, Dockery DW, Schwartz J, Zanobetti A. 2000b. The National Morbidity, Mortality, and Air Pollution Study, Part II: Morbidity and Mortality from Air Pollution in the United States. Research Report 94. Health Effects Institute, Cambridge MA.

Schwartz J, Zanobetti A, Bateson T. 2003. Morbidity and mortality among elderly residents of cities with daily PM measurements. In: Revised Analyses of Time-Series Studies of Air Pollution and Health, Special Report, pp 25–58. Health Effects Institute, Boston MA.



## RELATED HEI PUBLICATIONS: PARTICULATE MATTER

Number	Title	Principal Investigator	Date*
<b>Research Reports</b>			
124	Particulate Air Pollution and Nonfatal Cardiac Events <i>Part I: Air Pollution, Personal Activities, and Onset of Myocardial Infarction in a Case-Crossover Study</i> <i>Part II: Association of Air Pollution with Confirmed Arrhythmias Recorded by Implanted Defibrillators</i>	A Peters DW Dockery	2005 2005
99	A Case-Crossover Analysis of Fine Particulate Matter Air Pollution and Out-of-Hospital Sudden Cardiac Arrest	H Checkoway	2000
97	Identifying Subgroups of the General Population that May Be Susceptible to Short-Term Increases in Particulate Air Pollution: A Time-Series Study in Montreal, Quebec	MS Goldberg	2000
95	Association of Particulate Matter Components with Daily Mortality and Morbidity in Urban Populations	M Lippmann	2000
94	The National Morbidity, Mortality, and Air Pollution Study <i>Part I: Methods and Methodologic Issues</i> <i>Part II: Morbidity and Mortality from Air Pollution in the United States</i> <i>Part III: PM<sub>10</sub> Concentration-Response Curves and Thresholds for the 20 Largest US Cities</i>	JM Samet JM Samet JM Samet	2000 2000 2004
91	Mechanisms of Morbidity and Mortality from Exposure to Ambient Air Particles	JJ Godleski	2000
86	Statistical Methods for Epidemiologic Studies of the Health Effects of Air Pollution	W Navidi	1999
75	Ozone Exposure and Daily Mortality in Mexico City: A Time-Series Analysis	DP Loomis	1996
<b>Special Reports</b>			
	Revised Analyses of Time-Series Studies of Air Pollution and Health		2003
	Reanalysis of the Harvard Six Cities Study and American Cancer Society Study of Particulate Air Pollution and Mortality: A Special Report of the Institute's Particle Epidemiology Reanalysis Project		2000
	Particulate Air Pollution and Daily Mortality: The Phase I Report of the Particle Epidemiology Evaluation Project <i>Phase I.A: Replication and Validation of Selected Studies</i> <i>Phase I.B: Analyses of the Effects of Weather and Multiple Air Pollutants</i>		1995 1997

\* Reports published since 1990.

Copies of these reports can be obtained from the Health Effects Institute and many are available at [www.healtheffects.org](http://www.healtheffects.org).





## BOARD OF DIRECTORS

### **Richard F Celeste** *Chair*

President, Colorado College

### **Purnell W Choppin**

President Emeritus, Howard Hughes Medical Institute

### **Jared L Cohon**

President, Carnegie Mellon University

### **Gowher Rizvi**

Director, Ash Institute for Democratic Governance and Innovations, Harvard University

## HEALTH RESEARCH COMMITTEE

### **Mark J Utell** *Chair*

Professor of Medicine and Environmental Medicine, University of Rochester

### **Melvyn C Branch**

Joseph Negler Professor of Engineering, Mechanical Engineering Department, University of Colorado

### **Kenneth L Demerjian**

Professor and Director, Atmospheric Sciences Research Center, University at Albany, State University of New York

### **Peter B Farmer**

Professor of Biochemistry, Cancer Studies, and Molecular Medicine, Cancer Biomarkers and Prevention Group, University of Leicester

### **Helmut Greim**

Professor, Institute of Toxicology and Environmental Hygiene, Technical University of Munich

## HEALTH REVIEW COMMITTEE

### **Daniel C Tosteson** *Chair*

Professor of Cell Biology, Dean Emeritus, Harvard Medical School

### **Ross Anderson**

Professor and Head, Department of Public Health Sciences, St George's Hospital Medical School, London University

### **Ben Armstrong**

Reader in Epidemiological Statistics, Department of Public Health and Policy, London School of Hygiene and Public Health

### **John R Hoidal**

Professor of Medicine and Chief of Pulmonary/Critical Medicine, University of Utah

### **Thomas W Kensler**

Professor, Division of Toxicological Sciences, Department of Environmental Sciences, Johns Hopkins University

## OFFICERS & STAFF

**Daniel S Greenbaum** *President*

**Robert M O'Keefe** *Vice President*

**Jane Warren** *Director of Science*

**Jacqueline C Rutledge** *Director of Finance and Administration*

**Deneen Howell** *Corporate Secretary*

**Cristina I Cann** *Staff Scientist*

**Aaron J Cohen** *Principal Scientist*

**Maria G Costantini** *Principal Scientist*

**Wei Huang** *Staff Scientist*

**Debra A Kaden** *Principal Scientist*

### **Richard B Stewart**

University Professor, New York University School of Law, and Director, New York University Center on Environmental and Land Use Law

### **Robert M White**

President (Emeritus), National Academy of Engineering, and Senior Fellow, University Corporation for Atmospheric Research

### **Archibald Cox** *Founding Chair, 1980–2001*

### **Donald Kennedy** *Vice Chair Emeritus*

Editor-in-Chief, *Science*; President (Emeritus) and Bing Professor of Biological Sciences, Stanford University

### **Stephen I Rennard**

Larson Professor, Department of Internal Medicine, University of Nebraska Medical Center

### **Sylvia Richardson**

Professor of Biostatistics, Department of Epidemiology and Public Health, Imperial College School of Medicine

### **Howard Rockette**

Professor and Chair, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

### **Jonathan M Samet**

Professor and Chairman, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University

### **Ira Tager**

Professor of Epidemiology, School of Public Health, University of California, Berkeley

### **Brian Leaderer**

Professor, Department of Epidemiology and Public Health, Yale University School of Medicine

### **Edo D Pellizzari**

Vice President for Analytical and Chemical Sciences, Research Triangle Institute

### **Nancy Reid**

University Professor, Department of Biostatistics, University of Toronto

### **William N Rom**

Professor of Medicine and Environmental Medicine and Chief of Pulmonary and Critical Care Medicine, New York University Medical Center

### **Sverre Vedal**

Professor, Department of Environmental and Occupational Health Sciences, University of Washington

**Sumi Mehta** *Staff Scientist*

**Geoffrey H Sunshine** *Senior Scientist*

**Annemoo MM van Erp** *Senior Scientist*

**Terésa Fasulo** *Science Administration Manager*

**L Virgi Hepner** *Senior Science Editor*

**Jenny Lamont** *Science Editor*

**Francine Marmenout** *Senior Executive Assistant*

**Teresina McGuire** *Accounting Assistant*

**Kasey L Oliver** *Administrative Assistant*

**Robert A Shavers** *Operations Manager*



HEALTH  
EFFECTS  
INSTITUTE

Charlestown Navy Yard  
120 Second Avenue  
Boston MA 02129-4533 USA  
+1-617-886-9330  
[www.healtheffects.org](http://www.healtheffects.org)

RESEARCH  
REPORT

Number 94 Part IV  
September 2005