HEI

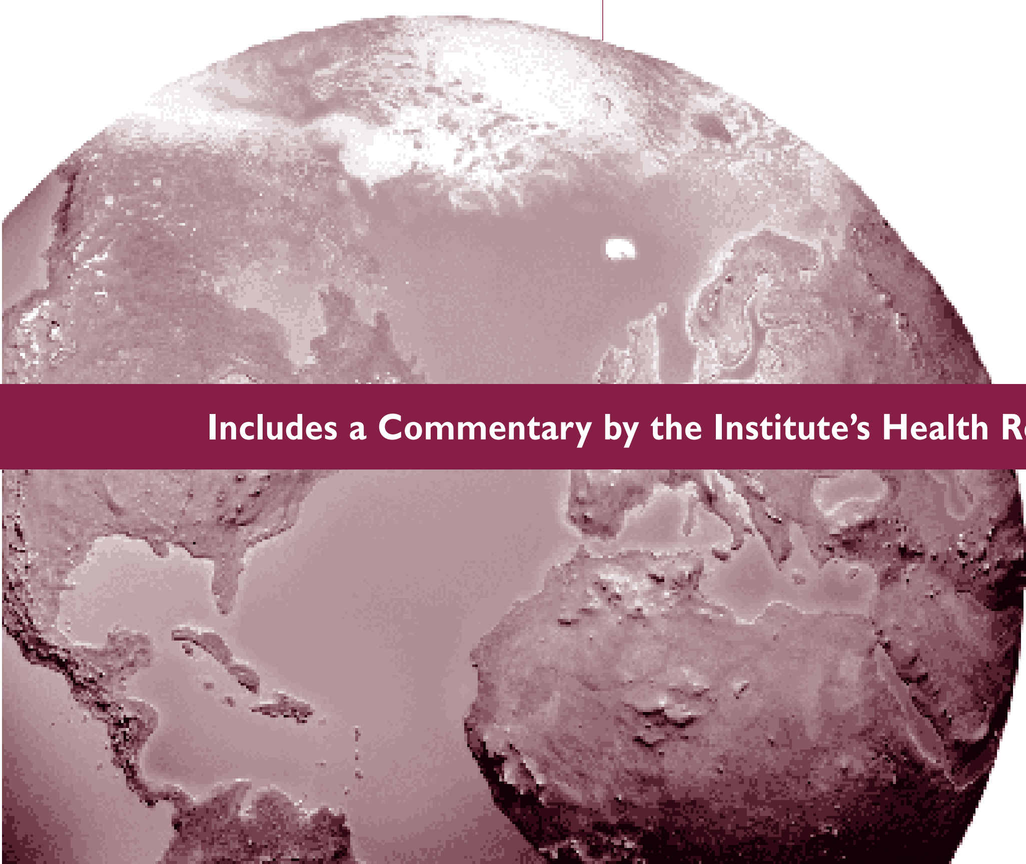
Walter A Rosenblith New Investigator Award
RESEARCH REPORT

HEALTH EFFECTS INSTITUTE

Number 123
December 2004

Time-Series Analysis of Air Pollution and Mortality: A Statistical Review

Francesca Dominici



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 123

Time-Series Analysis of Air Pollution and Mortality: A Statistical Review

BACKGROUND

In 1998, HEI issued the first Request for Applications (98-5) for the Walter A Rosenblith New Investigator Award, which had been established to provide funding for independent research by outstanding investigators at the beginning of their careers. Throughout his life, Professor Rosenblith had committed himself to developing young scientists. As the first Chair of the HEI Research Committee, he was instrumental in establishing HEI's scientific program and sought to attract basic scientists to air pollution research so the field may benefit from their new ideas and fresh approaches.

Dr Francesca Dominici was the first recipient of the Rosenblith Award in 2000. This Report reviews her collaboration with colleagues to develop innovative analytic methods for epidemiologic time-series studies.

Since the 1980s, many epidemiologic time-series studies have found associations between short-term increases of fairly low concentrations of ambient particulate matter (PM) and short-term increases in mortality and morbidity (eg, emergency room visits, hospitalizations). These studies were generally conducted in single locations chosen for a variety of reasons and were analyzed with a variety of different statistical approaches. Many people noted two fundamental questions about these studies: Could pollutants other than PM be responsible for these effects? and Would a study of many locations chosen on the basis of specific criteria and analyzed with uniform methods find similar associations? The National Morbidity, Mortality, and Air Pollution Study (NMMAPS), initiated in 1996 by HEI, was designed to address these and other questions raised about earlier studies.

Investigators at The Johns Hopkins University and Harvard University selected 90 US cities on the basis of population size and the availability of two types of data: PM_{10} mass concentrations and daily records of cause-specific deaths. The investigators developed new methods to evaluate the association between PM_{10} concentrations and mortality in the extensive data base and in subsets, depending on the type and quantity of data available for certain cities. They also evaluated other issues raised by earlier studies: (1) whether pollution data gathered by central monitors are an adequate surrogate for personal exposure (an assumption generally applied in such studies) and the degree to which measurement error affects the results obtained under such assumptions; (2) whether mortality was advanced by just a few days for frail, near-death individuals (mortality displacement) or by a longer time for other susceptible individuals; and (3) whether, in the PM concentration–mortality relation, a threshold could be found beneath which adverse events were not observed.

APPROACH

As one of the NMMAPS investigators, Dominici proposed to use funds from the New Investigator Award to develop and validate more flexible methods and statistical models to apply to the NMMAPS database to:

1. obtain national estimates of air pollution that would be resistant to mortality displacement and to biases from modeling long-term trends inappropriately;
2. determine the time course of health events after high concentrations of air pollutants (lagged effects) by assessing the short-, medium-, and long-term adverse health effects and testing the variability of these effects across locations and on different time scales;

Continued

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr Francesca Dominici at The Johns Hopkins University, Baltimore MD. The following Research Report contains a Preface, the detailed Investigator's Report and a Commentary on the study prepared by the Institute's Health Review Committee.

3. establish how the components of measurement error in exposure variables might influence risk assessment; and
4. evaluate the effects of air pollution on morbidity and mortality concurrently and on hospitalization rates for elderly residents of the 10 largest cities.

An important aspect of Dominici's work was her extensive investigation into the effects of model choices and assumptions on the results of data analysis. This led to discovering that part of a statistical software program that had been used extensively by many researchers in the field at that time was not, in its generally marketed format, entirely appropriate to analyze data in air pollution time-series studies. This discovery prompted Dominici to develop programming that would overcome the software's limitations.

COMMENT

The research conducted under this Award not only has advanced methods used to analyze complex time-series data, but also has highlighted the importance of undertaking in-depth sensitivity analyses, even when validated methods have been applied to the data. Fulfilling the goal of the Rosenblith Award, Dominici has developed from a promising researcher into a leader in the field of statistical methods for research on the health effects of air pollution.

 Cameographics, Union ME, Compositor. Printed at Capital City Press, Montpelier VT.
Library of Congress Catalog Number for the HEI Report Series: WA 754 R432.
The paper in this publication meets the minimum standard requirements of the ANSI Standard Z39.48-1984 (Permanence of Paper) effective with Report 21 in December 1988; and effective with Report 92 in 1999 the paper is recycled from at least 30% postconsumer waste with Reports 25, 26, 32, 51, 65 Parts IV, VIII, and IX, 91, 105, and 117 excepted. These excepted Reports are printed on acid-free coated paper.

HEALTH EFFECTS INSTITUTE

CONTENTS
Research Report 123

Time-Series Analysis of Air Pollution and Mortality: A Statistical Review

Francesca Dominici

Department of Biostatistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, Maryland

HEI STATEMENT

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

PREFACE

This Preface describes the HEI Walter A Rosenblith New Investigator Award: its intention, inception, and requirements.

INVESTIGATOR'S REPORT

When an HEI-funded study is completed, the investigator submits a final report. The Investigator's Report is first examined by three ouside 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 investigator has an opportunity to exchange comments with the Review Committee and, if necessary, revise the report.

COMMENTARY Health Review Committee

The Commentary about the Investigator's 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.

RELATED HEI PUBLICATIONS

Publishing history: This document was posted as a preprint on *www.healtheffects.org* and then finalized for print.

Citation for whole document:

Dominici F. December 2004. Time-Series Analysis of Air Pollution and Mortality: A Statistical Review. Research Report 123. Health Effects Institute, Boston MA.

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

PREFACE

This Research Report presents the accomplishments of a study funded under HEI's Walter A Rosenblith New Investigator Award. Throughout his career, Professor Rosenblith was well known for attracting and supporting people engaged in more basic scientific research who would bring new ideas and new tools to environmental questions. HEI established the Award in 1998 to provide funding for outstanding investigators who are beginning independent research. By providing financial support for investigators early in their careers, HEI hopes to encourage highly qualified individuals to undertake research on the health effects of air pollution.

Professor Rosenblith (1913–2002) served as the first Chair of HEI's Research Committee (from 1980 to 1989) and then as a member of the HEI Board of Directors (from 1990 to 1996). His vision of science and the standard of excellence he brought to all his endeavors enabled HEI to quickly develop a strong scientific research program. At his urging, the program would not only fund research to contribute needed scientific information for regulation, but also research to strengthen the fundamental science related to environmental issues.

The Request for Applications for the New Investigator Award is issued yearly. Currently, it provides up to $100,000 per year with a maximum of $300,000 for 3 years in total costs. The funds can be used to provide salary support for the investigator and supporting junior personnel, as well as operating costs that include supplies and equipment. The investigator is expected to devote at least 50% of his or her time to this project or related research. HEI expects to provide one or more awards from this RFA each year, depending on the number and quality of applications. Scientists of any nationality holding a PhD, ScD, MD, DVM, DrPH, or equivalent are eligible to apply. The candidates may have training and experience in any branch of science relevant to studying air pollution. The Research Committee considers three factors in evaluating applications: (1) the applicant's potential for becoming a leader in their chosen field of research; (2) evidence of tangible institutional support toward the candidate's professional development; and (3) the quality of the research proposed, especially if it will involve an innovative approach.

Francesca Dominici was the first scientist to receive the Walter A Rosenblith New Investigator Award. Her personal reflections about this award's importance to her and her work follow.

I remember vividly the day that Scott Zeger called to congratulate me for receiving the Walter A Rosenblith New Investigator Award from HEI. I was thrilled by this news, partly because having a statistical proposal receive the award indicated that the scientific community was highly valuing statistical contributions to air pollution research.

In this report, I summarize my contributions to time-series analyses of air pollution and mortality that were made possible with this support from HEI. These contributions are the result of truly unique relationships with my two mentors and friends, Jonathan M Samet and Scott L Zeger. Without their help and support, none of the work summarized in this report would exist. Jon taught me (1) how to write in English (an ongoing process); (2) the epidemiologic language (for example what epidemiologists call "effect modification" statisticians call "interaction"); and (3) how to communicate highly technical statistical concepts, like Bayesian hierarchical models, to nonstatisticians. Scott taught me (1) how to develop new statistical methods or best tailor existing methods to address important substantive questions; and (2) how to tackle the statistical analysis of messy data, such as time-series analyses of air pollution and mortality.

Several other people also contributed to this work in many different ways. In particular, I would like to thank Thomas Louis, who has always been available for advice and for constructive suggestions in the methodologic work for the National Morbidity, Mortality, and Air Pollution Study; Aidan McDermott, for his leadership in the development and update of the NMMAPS database; Aaron Cohen, who opened my eyes to the "real world" of air pollution research and who had the patience and the motivation to understand the technical aspects of my work; and Lianne Sheppard and Merlise Clyde, with whom I coauthored a review paper (see Dominici et al 2003d), portions of which are included in this Report.

In conjunction with an excellent research environment, The Walter A Rosenblith Award constitutes a very valuable opportunity for a junior investigator to further his or her research career. The substantial percentage of salary support covered by the award over a 3-year period, the highly valuable interaction with the HEI Research Committee (composed of internationally recognized scientists), and the interaction with the HEI Research Staff (which increases the impact of the work) are all targeted to promote the success of junior investigators. I strongly encourage other investigators to apply for this award.

INVESTIGATOR'S REPORT

Time-Series Analysis of Air Pollution and Mortality: A Statistical Review

Francesca Dominici

ABSTRACT

The Walter A Rosenblith New Investigator Award provided funding to explore new statistical approaches for air pollution research. This report reviews and summarizes the methodologic and substantive contributions to time-series analyses of air pollution and mortality that this award made possible. The review is organized according to the following general topics: (1) semiparametric methods for time-series analyses of air pollution and mortality; (2) explorations into the sensitivity of generalized additive models (GAMs*) applied to time-series data; (3) combining information in multisite time-series studies; (4) effects of misclassification of exposure; (5) mortality displacement; (6) shape of the concentration–response curve; and (7) ongoing projects and future directions. Appendix A includes abstracts of papers published as reports and in peer-reviewed journals.

INTRODUCTION

The statistical analysis of time-series data on air pollution and health presents several methodologic issues (US National Research Council 1998, 1999, 2001). Identifying the independent effects of a specific pollutant requires careful adjustment for simultaneous exposure to a complex mixture of copollutants. Extensive covariate adjustment is required to control for confounding factors (eg, age effects, weather variables, and seasonality). In addition, exposure measurement error can potentially lead to bias in estimates of the health effects of air pollution.

Because much of modern air pollution epidemiology is oriented toward the analysis and evaluation of complex data about the health effects of air pollution, advanced statistical methods have figured prominently in epidemiologic applications; such methods include generalized regression models for count and binary time-series data (Liang and Zeger 1986; McCullagh and Nelder 1989), GAMs (Hastie and Tibshirani 1990), and hierarchical models (Lindley and Smith 1972; Morris and Normand 1992).

In this report, I review statistical and substantive contributions to time-series analyses of air pollution and health outcomes, and discuss how these statistical methods have provided a clearer understanding of the extent to which air pollution affects human health. As part of the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) (Samet et al 2000a,b,c), our research team has assembled a large national database and produced:

- improved semiparametric regression models for time-series analyses of air pollution and health (Dominici et al 2004a);
- Bayesian hierarchical models for producing national, regional, and city estimates of the relative risk of mortality associated with concentrations of particulate matter 10 µm or smaller in aerodynamic diameter (PM_{10}) and other pollutants for the 88 largest urban centers in the United States (Dominici et al 2000a, 2002a, 2003a,b; Samet et al 2000a,b,c);
- spatial time-series models for creating maps of estimates for relative rates of mortality associated with shorter-term exposure to PM_{10} (Dominici et al 2003d);
- measurement error models for estimating the impact of exposure misclassification on time-series estimates of relative risks (Dominici et al 2000c; Zeger et al 2000);
- time-scale Poisson regression models for providing evidence in contradiction of the mortality displacement hypothesis (Zeger et al 1999; Dominici et al 2003c);
- Bayesian hierarchical models for combining dose-response curves and for providing evidence in contradiction of the

* A list of abbreviations and other terms appears at the end of the Investigator's Report.

This Investigator's Report is one part of Health Effects Institute Research Report 123, which also includes a Preface describing the Walter A Rosenblith New Investigator Award, a Commentary by the Health Review Committee, and an HEI Statement about the research project. Correspondence concerning the Investigator's Report may be addressed to Dr Francesca Dominici, Department of Biostatistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore MD 21205-3179; *fdominic@jhsph.edu*.

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.

threshold hypothesis (Daniels et al 2000; Dominici et al 2002a, 2003a,b; Daniels et al 2004b); and

- hierarchical bivariate time-series models for joint analysis of morbidity and mortality time-series data (Dominici et al 2004b).

Although progress has been made in investigating the association between air pollution and health, important questions still need to be addressed and considerable need remains for further data collection and methods development. At the end of this report, I provide an overview of our ongoing research projects to further advance our understanding of the health effects of air pollution.

SEMIPARAMETRIC METHODS FOR TIME-SERIES ANALYSES OF AIR POLLUTION AND MORTALITY

A widely used approach for time-series analyses of air pollution and health involves a semiparametric Poisson regression with daily mortality or morbidity counts as the outcome, linear terms measuring the percentage of increase in the mortality or morbidity associated with elevations in air pollution levels (the relative rates βs and smooth functions of time and weather variables to adjust for the time-varying confounders. Generalized linear models with parametric splines (eg, natural cubic splines) (McCullagh and Nelder 1989) or GAMs with nonparametric splines (eg, smoothing splines or locally weighted smoothers [LOESS]) (Hastie and Tibshirani 1990) are used to estimate effects associated with exposure to air pollution while accounting for smooth fluctuations in mortality that confound estimates of the effects of pollution. These statistical models take the following additive form:

$$E(Y_t) = \exp\{\beta_0 + \beta X_{t-\ell} + S(\text{time}, \lambda_1) + S(\text{temp}, \lambda_2) + \gamma \times \text{DOW}\} \quad (1)$$

where Y_t are daily mortality (or morbidity counts), X_t are daily levels of ambient air pollution levels, ℓ is the lag time of the pollution exposure (which is generally restricted to between 0 and 7 days), temp is a measure of temperature (average daily or dew point or both), and γ is the vector of regression coefficients associated with indicator variables for day of the week (DOW). The function $S(\cdot, \lambda)$ denotes a smooth function of a covariate (calendar time, temperature, humidity). The model is often constructed using smoothing splines, LOESS, or natural cubic splines with a smoothing parameter, λ, which represents the number of degrees of freedom in the smoothing spline, the span in the LOESS, or $\lambda - 2$ interior knots in the natural cubic splines. The parameter of interest, β, describes the change in the logarithm of the population average mortality count per unit of change in $X_{t-\ell}$, and it is generally interpreted as the percentage of increase in mortality for every 10 units of increase in ambient air pollution levels at lag ℓ.

Rather than focus on the effects associated with a single lag of a pollution variable, distributed lag models and time-scale models can be used to estimate cumulative and longer time-scale health effects of air pollution. Distributed lag models (Almon 1965; Zanobetti et al 2000b, 2002) are used to estimate associations between health outcomes on a given day and air pollution levels several days earlier by replacing $\beta X_{t-\ell}$ in model (1) with

$$\theta \sum_{\ell=1}^{L} \eta_\ell X_{t-\ell} \text{ and } \sum_{\ell=1}^{L} \eta_\ell = 1,$$

where θ measures the cumulative effect, and η_ℓ measures the contribution of the lagged exposure $X_{t-\ell}$ to the estimation of θ. Time-scale models (Zeger et al 1999; Schwartz 2000b, 2001; Dominici et al 2003d) are used to estimate associations between smooth variations of air pollution and daily health outcomes by replacing $\beta X_{t-\ell}$ in model (1) with

$$\sum_{k=1}^{K} \beta_k W_{kt},$$

where $W_{1t}, \ldots, W_{kt}, \ldots, W_{Kt}$ is a set of orthogonal predictors obtained by applying a Fourier decomposition to X_t, such that

$$\sum_{k=1}^{K} W_{kt} = X_t.$$

The parameters β_k denote the log relative rate of the health outcome for increases in air pollution at time scale k. Time scales of interest are short-term air pollution variations (1 to 4 days) and longer-term variations (1 to 2 months), which capture acute and less-acute health effects. Beyond 2 months, it is likely that any effects are dominated by seasonal confounding.

EXPLORATION OF GAM APPLICATIONS TO TIME-SERIES DATA

In 2002, the US Environmental Protection Agency (EPA) was finalizing its review of the evidence on particulate air pollution. Our team discovered that implementing GAMs to analyze time-series data of air pollution and health outcomes using the *gam* function of the S-Plus statistical software (Insightful Corp, Seattle WA) could overestimate the air pollution effects. More specifically, in these applications, the original default parameters of the *gam* function in S-Plus were found inadequate to guarantee the convergence of the backfitting algorithm (Dominici et al 2002b). In addition, Canadian investigators pointed out that the *gam* function in S-Plus, as programmed at that time, calculated the standard errors of the linear terms (the air pollution coefficients) by approximating the smooth terms as linear functions; this resulted in an underestimation of uncertainty (Chambers and Hastie 1992; Klein et al 2002; Lumley and Sheppard 2003; Ramsay et al 2003; Samet et al 2003).

Computational and methodologic concerns in implementing GAMs for time-series analyses of pollution and health delayed the review of the Criteria Document for PM because the time-series findings were a critical component of the evidence. The EPA deemed it necessary to reevaluate the results obtained from GAM analyses in time-series studies that had been considered to be key studies in the decision-making process. EPA officials identified nearly 40 published articles and requested that the investigators conduct new analyses of their data by applying (1) the *gam* function of S-Plus with appropriate programming changes, and (2) other methods that would estimate standard errors more accurately. These new analyses were peer reviewed by Special Panels of the HEI Health Review Committee that included epidemiologists and statisticians. Results of the new analyses and Commentaries by the Special Panels have been published in an HEI Special Report (Health Effects Institute 2003).

These new analyses of time-series studies highlighted a second important epidemiologic and statistical issue known as confounding bias. The relative rate estimates for the effect of pollution on mortality or morbidity could be influenced by observed and unobserved time-varying confounders (such as weather variables, season, and influenza epidemics) that vary in a similar manner as the time series data for air pollution and mortality or morbidity. To control for confounding bias, smooth functions of time and temperature variables are routinely included in the semiparametric Poisson regression model.

Adjusting for confounding bias is a more complicated issue than properly estimating the standard errors of the air pollution coefficients. The degree of adjustment for confounding factors, which is controlled by the number of degrees of freedom in the smooth functions of time and temperature, can have a large impact on the magnitude and statistical uncertainty of the relative rate estimates for mortality or morbidity. In the absence of a strong biological hypothesis, the choice of the number of degrees of freedom has been based on either expert judgment (Kelsall et al 1997; Dominici et al 2000a) or on optimality criteria; these include the minimum prediction error (based on the Akaike information criterion) and the minimum sum of the absolute value of the partial autocorrelation function of the residuals (Toulomi et al 1997; Burnett et al 2001).

Motivated by these observations, we improved the semiparametric regression approach for risk estimation in time-series analyses by:

- calculating a closed-form estimate of the asymptotically exact covariance matrix of the linear component of a GAM (the air pollution coefficients); to ease the implementation of these calculations, we developed GAM.EXACT, an extended version of the *gam* function of the S-Plus program (software available at *http://ihapss.biostat.jhsph.edu/software/gam.exact/gam.exact.htm*); (our GAM.EXACT software improved estimating the statistical uncertainty of the air pollution risk estimates); and
- developing a bandwidth selection strategy for the smooth functions of time and temperature that leads to air pollution risk estimates with small confounding biases shown relative to the standard errors of the estimates; (our bandwidth selection method was applied to four NMMAPS cities with daily air pollution data and compared with other bandwidth selection methods frequently used in such studies (Dominici et al 2004a).

By allowing a more robust assessment of all sources of uncertainty in air pollution risk estimates, including standard error estimation, confounding bias, and sensitivity to the choice of model, the application of our methods enhances the credibility of all time-series studies that are pertinent to the current policy debates about pollutant standards.

COMBINING INFORMATION IN MULTISITE TIME-SERIES STUDIES

In the past, single-site time-series studies have been criticized because they did not adequately represent the study locations and the heterogeneity of the statistical approaches used to estimate the associations between air pollution and health (Li and Roth 1995; Lipfert and Wyzga 1995). These criticisms have been addressed by using multisite studies in which site-specific data on air pollution and health were assembled under a common framework and analyzed with an uniform analytic approach: either fitting site-specific models to each site (Katsouyanni et al 1997); or applying the same model to all sites (the same variables and smoothing functions for potential confounding factors, although parameter estimates and fitted smooth functions were allowed to vary from site to site) (Samet et al 2000d). Hierarchical models have provided an appropriate approach for summarizing and integrating the findings of research studies in a particular geographical area (Lindley and Smith 1972; Morris and Normand 1992; Gelman et al 1995; Carlin and Louis 1996). Hierarchical models have been used by statisticians for the last four decades; because computational tools that facilitate their implementation have been developed (Thomas et al 1992; Gilks et al 1996), they have been applied widely in many disciplines, including analysis of multisite time-series data (Burnett and Krewski 1994; Katsouyanni et al 1997; Roemer et al 1998; Dominici et al 2000a; Schwartz 2000a; Zanobetti et al 2000a).

Bayesian hierarchical modeling is an appropriate and unified approach for combining evidence across studies because it quantifies the sources of variability and identifies effect modification (see Dominici [2002] for a more detailed discussion about using hierarchical models in multisite time-series studies). For example, we can assume a two-stage hierarchical model with the following structure: (I) Given multiple sites with time-series data for air pollution and daily mortality or morbidity counts, the association between air pollution and health is described using the site-specific regression model (1), which takes into account potential confounding factors such as trend, season, and climate. (II) The information from multiple sites is combined in a linear regression model in which the outcome variable (β^s) is the true relative mortality rate associated with site-specific air pollution indices within each site, and the explanatory variables $\left(X_j^s\right)$ are site-specific characteristics (population density, yearly averages of the pollutants, and temperature). Formally:

$$\beta^s = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_j^s + \text{ error.} \qquad (2)$$

If the explanatory variables, or predictors, X_j^s and j=1,..., p are centered around their means, the intercept (α_0) can be interpreted as the pooled effect for a site with mean predictors. The regression parameters (α_j) measure the change in the true relative rate of mortality associated with a unit of change in the corresponding site-specific variable.

The sources of variation in the estimation of health effects of air pollution are specified by the levels of the hierarchical model. The variation of $\hat{\beta}^s$ around β^s is described by the within-site variance (v^s), which depends on the number of days with available air pollution data and on the predictive power of the site-specific regression model. The variation of β^s around α_0 is described by the between-site variance (τ^2), which measures the heterogeneity of the true air pollution effects across cities. The specification of a Bayesian hierarchical model is completed with the selection of the prior distributions for the parameters at the top level of the hierarchy. If there is no desire to incorporate prior information into the analysis, the default choice is to use conjugate prior distributions with large variances. However, it is important to complete the Bayesian analysis by investigating the sensitivity of the substantive findings to the prior distributions.

Posterior distributions of the pooled estimate α_0, of the between-site variance τ^2, and of the second-stage regression parameters α_j provide an overall summary of the site-specific relative rates of mortality, a characterization of the heterogeneity of the air pollution effects across the several locations, and the identification of site-specific characteristics that modify the association between air pollution and health. The two-stage hierarchical approach thus described can be extended to include additional levels of the hierarchical models (for example, sites within geographical regions, geographical regions within nations, etc) that lead to the estimation of additional sources of variability (within-site, between-site within region, and between regions) and to potential effect modifiers at the site or regional level (see, for example, Dominici et al 2002a).

Complex hierarchical models can be fitted using simulation-based methods (Tierney 1994; Gilks et al 1996) that provide samples from the posterior distributions of all parameters. Alternatively, a point estimate of the pooled effect can be obtained by assuming a random effects model and by taking a weighted average of the site-specific estimates (as, for example, suggested by DerSimonian and Laird 1986). Under the weighted average approach for a random effects model, the weights of the site-specific estimates are modified to take into account the variability between locations (eg, by including a point estimate of τ^2). A Bayesian approach more completely assesses the heterogeneity of the effects across locations because inspecting the posterior distribution of τ^2 provides a better characterization of the degree of heterogeneity of the effects across sites than does a point estimate of τ^2 or the classical χ^2 test of $\tau^2 = 0$.

We estimated city-specific, regional, and national air pollution effects by applying a three-stage hierarchical model to the 88 largest metropolitan areas in the United

States that had data from 1987 through 1994 (Dominici et al 2002a). The pooled regional estimates of the PM_{10} effects varied somewhat across the regions; they were estimated to be greatest in the Northeast, where we calculated a relative rate of 0.41% increase in mortality per 10-μg/m^3 increase in PM_{10} (95% posterior interval 0.04, 0.78). The national average relative rate was 0.21% increase in mortality per 10-μg/m^3 increase in PM_{10} (95% posterior interval 0.09, 0.33) (Dominici et al 2003b).

EFFECTS OF MISCLASSIFICATION OF EXPOSURE

One barrier to interpreting observational evidence about the adverse health effects of air pollution is the measurement error inherent in estimates of exposure based on ambient pollutant monitors. Exposure assessment studies have shown that data from monitors at central sites do not adequately represent personal exposure (Lioy et al 1990; Mage and Buckley 1995; Ozkaynak et al 1996; Janssen et al 1997, 1998; Haran et al 2002). Thus, the exposure error that results from using centrally measured data as a surrogate for personal exposure can potentially lead to bias in estimates of the health effects of air pollution (Thomas et al 1993). Nevertheless, because regulations are based on ambient air and most epidemiologic studies rely on pollutant measurements from central-site monitors to estimate exposure, many epidemiologic studies assess health effects on the basis of ambient measurements regardless of their intention to evaluate the effects on the basis of actual personal exposure.

Both ambient concentrations and personal air pollutant exposures vary over time. In addition, personal exposures vary substantially for different individuals. The microenvironments where individuals spend their time have different dominant sources of pollution, all of which additively contribute to total personal exposure. Microenvironmental modeling is a popular approach for exposure assessment, particularly when directly measuring total personal exposure is not possible (Duan 1991; Rodes et al 2001).

Individual personal exposure can be partitioned into ambient (outdoor) versus nonambient (nonoutdoor) sources, of which personal exposure to ambient sources is of interest from a regulatory perspective. A simple model for total (P) personal exposure (X) for individual i at time t, X_{it}^P has the form

$$X_{it}^P = X_{it}^N + \alpha_{it} X_{it}^A$$

where X_{it}^N is personal exposure from nonambient (N; nonoutdoor) sources, X_{it}^A is the ambient (A; outdoor) concentration at individual i's spatial location, and α_{it} is an attenuation parameter defined as the fraction of the ambient concentration that an individual experiences as exposure (Ott et al 2000; Sheppard and Damian 2000; Wilson et al 2000). The attenuation parameter depends upon the penetration, deposition, and decay rates for a specific pollutant in the microenvironment, as well as on individual behavior (particularly how much time is spent in microenvironments).

Current research suggests that, for large populations, it may be reasonable to assume that $\alpha_{it} = \alpha$, that $X_{it}^A \propto X_t^A$, and that ambient and nonambient sources are independent. This has important consequences for study design, particularly because most studies rely on the measurement $\widehat{X}_t^A$ of X_t^A from a fixed-site ambient monitor.

In considering the consequences for estimating health effects of air pollution by using surrogate measures of exposures, one approach (Zeger et al 2000) begins by decomposing the difference between pollution measurements X_{it}^P and $\widehat{X}_t^A$ into three components:

$$X_{it}^P = \widehat{X}_t^A + \left(X_{it}^P - \overline{X}_t^P\right) + \left(\overline{X}_t^P - X_t^A\right) + \left(X_t^A - \widehat{X}_t^A\right) \tag{3}$$

where $\left(X_{it}^P - \overline{X}_t^P\right)$ is the error due to having used aggregated data rather than individual exposure data (or [total personal exposure] − [mean aggregated personal exposure]); $\left(\overline{X}_t^P - X_t^A\right)$ is the difference between the mean aggregated personal exposure and the true ambient pollutant concentration; and $\left(X_t^A - \widehat{X}_t^A\right)$ is the difference between the true ambient pollutant concentration and the measured ambient pollutant concentration. Regardless of the epidemiologic design, each of these differences may introduce some degree of exposure measurement error.

For example, in an ecological time-series study of air pollution and health, suppose that we are interested in drawing inferences about the relation between Y_t and $\overline{X}_t^P$; we have substantial information about the relation between Y_t and $\widehat{X}_t^A$ at a particular site and the results of separate studies of the association between $\widehat{X}_t^A$ and $\overline{X}_t^P$. Here we assume $\widehat{X}_t^A - X_t^A$. In this scenario, a reasonable approach is to build a two-stage model. At the first stage, we assume:

$$E\left(Y_t\right) = \exp\left(\overline{X}_t^P \beta_P \ + \ \text{confounders}\right) \tag{4}$$

where β_P is the log relative rate of death associated with a unit change in average personal exposure, $\overline{X}_t^P$ is (in this example) a missing covariate, and the confounders are the same as in model (1). At the second stage, we assume a measurement error model for the relation of average personal exposures and ambient concentrations, taking account of variation within and across locations. This modeling approach combines a log-linear model for Y_t given $\overline{X}_t^P$ with measurement error model for $\overline{X}_t^P$ given X_t^A to make inference about β_P.

As discussed below, the direct regression of Y_t on X_t^A, given β, is also of interest from a regulatory perspective because only ambient (outdoor) concentrations of pollutants are currently regulated. This modeling approach (Dominici et al 2000c) can be described as a combination of Bayesian hierarchical modeling (Lindley and Smith 1972; Morris and Normand 1992) and data augmentation (Tanner 1991); it is an example of regression calibration, which is widely used for handling measurement error in nonlinear models (Carroll et al 1995).

Regardless of the sampling design, studies of air pollution and health are subject to the limitations posed by the available air pollution measurements. Various aspects of air pollution exposure measurements, such as the three exposure components described in equation (3), produce different types of bias in the health effect estimates; such biases carry more or less importance in different sampling designs. Furthermore, reexamining simplified assumptions will be needed as our understanding of exposure measurement distributions develops. A systematic treatment of exposure measurement error and modeling issues for different sampling designs, including recommended data collection and novel statistical methods, is still lacking in the air pollution literature.

MORTALITY DISPLACEMENT

Even if the associations between particulate air concentrations and mortality are causal, it is still important to know who is affected and by how much: if only those very near death are susceptible, then the public health importance of particulate air pollution might still be fairly small. Short-term mortality displacement describes a model wherein a pool of frail, near-death individuals is identified; the pool remains relatively stationary due to the effects of life-threatening illnesses, other diseases such as pneumonia, and conditions such as congestive heart failure. These individuals have a short life expectancy; the effect of high air pollution concentrations is to hasten their deaths by a matter of days. This effect acts to somewhat deplete the pool of frail individuals so that the higher mortality rate immediately after a day of high air pollution is followed by a lower mortality rate for the next few days. Fundamental to this model is the assumption that particulate air pollution does not affect the rate at which people enter the pool of frail individuals.

Several statistical approaches have been developed to investigate short-term mortality displacement. Poisson regression methods in the frequency domain (Kelsall et al 1999; Zeger et al 1999; Schwartz 2001) and in the time domain (Schwartz 2001; Dominici et al 2003d) estimate associations between air pollution mortality at different time scales of variation in both exposure concentrations and outcomes. These approaches can be used to test the mortality displacement hypothesis: if the association between air pollution and mortality is evident at only the shortest time scales, then pollution-related deaths are advanced by only a few days and therefore the theory of mortality displacement is supported.

Findings from recent time-scale analyses are inconsistent, however, with the hypothesis of short-term mortality displacement (Zeger et al 1999; Schwartz 2001). In fact, they suggest that smoother variations in the air pollution time series (14 days to 3 days) tend to be more strongly associated with mortality than less smooth variations (less than 2 days).

Although they are innovative, frequency-domain and time-scale approaches have limitations and their findings need to be interpreted with caution. Relative rates of mortality associated with long time-scale variations in air pollution (say longer than 2 months) tend to be heavily confounded by seasonality; interpretation of these time-scale relative rates is not straightforward. Distributed lag models (Zanobetti et al 2000b) are reasonable alternatives and allow a better interpretation of the coefficients, but are still affected by long-term confounding.

SHAPE OF THE CONCENTRATION–RESPONSE CURVE

One approach for estimating the relation between air pollution concentrations and daily deaths or hospital admissions (the concentration–response curve) in time-series studies is to model the logarithm of the expected value of daily mortality as a smooth function of air pollution after adjusting for other confounders:

$$E(Y_t) = \exp\left[S(X_t, \text{knots} = \nu) + \text{confounders}\right] \quad (5)$$

where $S(\cdot,\cdot)$ is modeled as a natural cubic spline with k knots at locations $\nu = (\nu_1, \ldots, \nu_k)$, and confounders include smooth functions of time and temperature, as described in model (1). The number and locations of knots are generally fixed in advance (as was done by Daniels and colleagues [2000] in the analysis of the 20 largest US cities), or can be estimated from the data using the Reversible Jump Monte Carlo Markov Chain (RJMCMC) (Green 1995) (as was used by Dominici and colleagues [2002a] in an analysis of the 88 largest US cities). Other methods can be used as alternatives to the natural cubic spline, such as smoothing splines, LOESS, or B-splines (Schwartz 2000a; Schwartz and Zanobetti 2000; Smith et al 2000). Shapes of concentration–response curves are generally not very sensitive to the smoothing method (except at extremes), but can be sensitive to the smoothness penalty or the number of knots.

For a fixed number of knots, the concentration–response model (5) can be expressed as a parametric model

$$E(Y_t) = \exp(\beta B_t + \text{confounders}) \tag{6}$$

where B_t is the t-row of the design matrix for the natural cubic spline with knots at locations ν of X_t, and β is the corresponding vector of coefficients.

To examine the question of whether the health effects of air pollution are not measurable below some level, a linear threshold model can be used:

$$E(Y_t) = \exp\left[\theta(X_{t-\ell} - h)^+ + \text{confounders}\right] \tag{7}$$

where the $X_{t-\ell}$ term in model (1) is replaced with a term of the form $(X_{t-\ell} - h)^+$, where $(x)^+ = x$ if $x \geq 0$ or $(x)^+ = 0$ if $x < 0$, and h is an unknown change-point (or threshold) that is estimated from the data. Bayesian methods can be used to estimate marginal posterior distribution of the location of the threshold h (Cakmak et al 1999; Smith et al 2000; Daniels et al 2000; Schwartz 2000a). The parameters β in the generalized additive model (1) and θ in model (7) have different interpretations: β measures the percentage of increase in mortality per unit of increase in air pollution at any level of the pollutant; θ measures the percentage of increase in mortality per unit of increase in air pollution when the pollutant level is higher than h. For pollutant values smaller than h, the log relative rate β is set to zero.

In multisite time-series studies, it is of interest to estimate a pooled concentration–response relation by combining concentration–response curves across locations. The hierarchical approach described earlier can be used here if equation (2) is generalized to the multivariate case as

$$\beta^s = \alpha_0 + N(0, \Sigma) \tag{8}$$

where β^s and α_0 are the city-specific coefficient and the overall vector of the spline coefficients, respectively; and Σ denotes the between-city covariance matrix of the β^c. Effect modification of the shape of the dose-response curve also can be investigated by including city-specific covariates in equation (8).

Dominici and associates (2002a) recently used the hierarchical spline model defined in equations (6) and (8) with unknown numbers and locations of knots to estimate regional and national concentration–response relations in the NMMAPS database. The national concentration–response curve, obtained by combining information across all the cities, is clearly linear, and the posterior probability of zero knots from the RJMCMC is almost 1. These findings are consistent with previous analyses of time-series data in a number of locations including London, Cincinnati, Birmingham, Utah Valley, and Shenyang (Pope 2000), and with findings of the two major long-term studies, the Harvard Six Cities and the American Cancer Society cohort studies (Dockery et al 1993; Pope et al 1995).

The consistent finding of a linear relation between pollutant concentrations and adverse health outcomes, now confirmed at the national level, places a difficult burden on policy-makers; they are charged by the Clean Air Act to set protective standards for the public's health, and those standards must include a margin of safety for known susceptible populations. However, this linear relation might also reflect that data outside the linearity have been averaged out at regional and city levels, and for susceptible subgroups of the population (eg, elder individuals or people with preexisting diseases). On the other hand, substantial statistical power would be needed to provide strong evidence of nonlinearity in concentration–response relations for very specific strata of geographical and population subgroups.

These methods provide evidence in a form that would be directly applicable to policy development, as shown by our finding that the concentration–response relation for air pollution and mortality is linear with a high degree of certainty. With repeated application, our methods would offer an approach for tracking the health effects of air pollution over time as control measures are implemented; our methods should also have applications to other environmental health problems.

ONGOING PROJECTS AND FUTURE DIRECTIONS

During the last few years, a large number of epidemiologic studies have provided important evidence on the health effects of air pollution, and several state-of-the-art statistical approaches for analyzing environmental data have been proposed. Nevertheless, considerable need remains for a multidisciplinary approach that would support more effective risk analysis. As a result of the work summarized in this report, our team at The Johns Hopkins University has learned many important lessons that have enabled us to build a broader research program in air pollution research. Some of the program's highlights are discussed below.

Research in this field would be greatly enhanced by a systematic integration of research efforts from the following fields to establish an ongoing surveillance and accountability system; that is, a system to monitor pollutant concentrations, sources, and health effects, and also evaluate the role of new regulations in improving health status in populations by curtailing air pollution.

- ***Computer science and database management***: to integrate and systematically update national databases on air pollution, weather, health outcomes, individual-level risk factors, and area-level characteristics;
- ***Exposure assessment***: to continue characterizing air pollution exposure (especially a better understanding of the toxicity of particles), exposure misclassification (such as indoor versus outdoor exposures), and the effects of population mobility on exposure assessment;
- ***Environmental epidemiology***: to formulate and evaluate hypotheses about possible biologic pathways that may help explain the effects of short- and long-term exposure to air pollution;
- ***Social epidemiology***: to formulate hypotheses on interactions among socioeconomic, behavioral, and environmental determinants of disease outcomes;
- ***Biostatistics***: to develop statistical methods to analyze highly complex data structures and to estimate health risk by taking into account all sources of uncertainty;
- ***Health economics***: to develop methods for estimating medical expenditures attributable to diseases related to air pollution exposure and to carry out related analyses;
- ***Policy***: to translate relative risk estimates (such as calculating the expected loss of life from various air pollution scenarios) into specific measures that will be relevant to establishing public policies.

As a first step toward an integrated approach to air pollution research, we have assembled a multidisciplinary team of investigators called the Environmental Biostatistics and Epidemiology Group. The Group's efforts are aimed at promoting research and training at the various interfaces between the disciplines listed above. Regular participants of the Group are faculty members, post-doctoral fellows, and students in the Departments of Biostatistics, Epidemiology, Environmental Health Sciences, and Health Policy and Management at Johns Hopkins, plus consultants and faculty members from other institutions. Below are descriptions of some of our data resources and ongoing projects; more details, including a list of participants and a more complete description of our research, are given at our website *www.biostat.jhsph.edu/bstproj/ebeg/index.php*.

COMPUTER SCIENCE AND DATABASE DEVELOPMENT

Internet-Based Health and Air Pollution Surveillance System: A New Model for Health Monitoring*

The purpose of this research program is to create an internet-based system for monitoring the effects of air pollution on mortality and morbidity in US cities. The association between shorter-term fluctuations in concentrations of particles and other air pollutants with daily mortality and morbidity is important evidence when the EPA and other organizations deliberate air quality standards. Key constituents in these deliberations (including government, industry, nongovernmental organizations, and the public) depend upon reliable, up-to-date information about pollution-associated health effects. The data necessary for estimating these effects, including those for mortality and morbidity, air pollution, weather, and demographics, are routinely collected by several government agencies at enormous expense to the public. However, information that is relevant to environmental policy can only be produced through systematic analysis of these data. Knowledge is gained by further interpreting the detailed results of many studies and approaches. Among the analyses of public databases that have been performed and reported is the HEI-initiated NMMAPS, which forms the basis of this Investigator's Report.

It is the results of the NMMAPS research project that have led us to propose creating a web-based system that regularly accesses, analyzes, and disseminates policy-relevant data about the association between air pollution and mortality and morbidity in US cities. In Phase I of this project, the requisite public information and results of statistical analyses will be posted to an Internet-Based Health and Air Pollution Surveillance System website so that all constituents can stay informed about the most recently available data and results for each of the System's components. In Phase II, to be undertaken soon, we are planning to create web-based methods for easily conducting new, innovative analyses.

* Project funded by HEI. The principal investigator is Scott L Zeger and the project leader is Aidan McDermott. More information is available at *http://ihapss.biostat.jhsph.edu*.

Hopkins Integrated National Tools†

As part of our ongoing studies on pollution and health, we are creating the Hopkins Integrated National Tools, made up of the following databases linked by zip code of residence.

National Air Pollution Monitoring Network (1987–2000): comprises daily values for PM_{10}, $PM_{2.5}$, ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2); data have been collected from approximately 3000 monitoring stations in the United States.

National Weather Monitoring Network (1987–2000): comprises average daily temperature and average dew point temperature from approximately 8000 monitoring stations in the United States.

National Medicare Cohort (1999–2002): comprises individual-level longitudinal data ("time to event" and "time of entry into the cohort") for prescription drugs, hospitalizations, and deaths for all Medicare participants (almost all individuals in the United States who were recorded as being older than 65 years in the 2000 US census). It also includes individual-level information on age, gender, and ethnicity. In 2000, this database contained information on approximately 48 million Medicare participants, 8 million deaths, and 15 million hospitalization records.

The Medicare Current Beneficiary Survey (1999–2002): comprises a nationally representative subcohort of 13,000 Medicare beneficiaries. In addition to the Medicare administrative information, the Medicare Current Beneficiary Survey participants answered questions related to (1) income, household size, marital status, education; (2) daily activities, behaviors that affect health [smoking, exercise]; (3) hearing and vision, (4) disabilities and impairments; (5) height and weight; (6) general health and recent changes in health; (7) history and current status of a number of medical conditions (including cancer, heart disease, emphysema, asthma, pneumonia, and respiratory infections); (8) procedures that could affect health and health care, such as smoking, flu shots, pap smears, and mammograms; (9) use of managed care medical services under Medicare; and (10) use of non-Medicare medical services (for example, at a Veterans' Administration medical facility).

US Census (2000): comprises census data at the level of zip code for community characteristics. For residents, these include socioeconomic status (eg, household income, income below poverty, highest education attained), transportation (eg, type of transportation used to get to work), and employment (eg, industry type, occupation); for communities, these include the number of industrial facilities (eg, waste treatment and disposal, power generation) and health service organizations (eg, ambulance services and hospitals).

† Project funded by the EPA as part of the National Medicare Pollution Study. The project leader is Aidan McDermott.

EPIDEMIOLOGIC STUDIES

The National Medicare Pollution Study‡

The National Medicare Pollution study is the first large epidemiologic study on pollution and health that will use a government-based national data source to: (1) routinely monitor the health of the population; and (2) directly inform environmental policy on the acute and chronic health effects associated with short-term and long-term exposure to air pollution. Our Medicare cohort includes everyone in the United States enrolled in the Medicare system in 1998 (approximately 48 million people, mostly older than 65) for whom 4 years of follow-up information was available about individual-level medical information, such as diagnoses and medical bills. As part of the Hopkins Integrated National Tools program, each individual-level record is linked by place of residence (zip code) to the closest air pollution monitor; this leads to approximately 18 million individual-level records of $PM_{2.5}$ exposure levels and data for approximately 800,000 deaths.

Under the National Medicare Pollution study we plan to evaluate the following set of issues: (1) the associations between both long-term and short-term air pollution exposure and the risk of death; (2) the associations between both long-term and short-term air pollution exposure and the risk of disease; and (3) possible differential effects between long-term and short-term exposure within separate "high-risk" cohorts characterized by certain conditions that enhance their susceptibility to the effects of air pollution exposure.

A unique feature of this design is that it allows us to estimate both acute and chronic health effects within the same population that are associated with short-term and long-term changes in pollution levels. Therefore, we plan to identify the contributions of short- and long-term pollution exposure to acute and chronic health effects; this assessment will provide critical information on possible biological mechanisms that might underlie the observed associations between air pollution and adverse health effects.

‡ Project funded by the EPA. The principal investigator is Jonathan M Samet.

Air Pollution and Health: A European and North American Approach**

Multicity studies of air pollution, daily mortality, and hospital admissions have now been carried out in Europe, the United States, and Canada. They provide strong evidence that exposure to particulate air pollution increases rates of morbidity and mortality from cardiovascular and respiratory diseases. They also suggest that these effects vary in magnitude across Europe and North America.

As an appropriate follow-up to these multicity studies, Air Pollution and Health: A European and North American Approach will bring together the investigators who have carried out two studies under the title Air Pollution and Health: A European Approach, NMMAPS in the United States, and several studies supported by Health Canada.

The overall goal is to use a common analytic framework to characterize the effects of air pollution on mortality and morbidity in Europe and North America, including the evaluation of spatial variation in health effects. The goal will be accomplished first by (1) creating an approach to establish the comparability of methods used by the different investigative groups; (2) developing and applying analytic methods for characterizing the heterogeneity of air pollution effects across locations, and (3) exploring methods to evaluate the degree of mortality displacement.

Once the methods are developed, they will be applied to a joint and parallel analysis of the air pollution and health data from existing databases on mortality and hospitalization from the three collaborating groups. In total, these databases cover many of the major time-series studies from developed countries that have been reported over the last decade.

BIOSTATISTICAL METHODS FOR ENVIRONMENTAL EPIDEMIOLOGY††

Evidence from environmental epidemiologic research often contributes to the foundation of major policy decisions, driving policy makers to pose challenging questions to researchers. These questions are often best answered by using statistical methods that characterize the risk of a targeted environmental agent while taking into account other environmental variables. The nature and characteristics of environmental data and health outcomes make the risk estimation challenging and require novel statistical methods to be developed.

The purpose of this research is to develop models to conduct integrated analyses of spatiotemporal data on exposure, health outcomes, and covariates; these data have been incompletely recorded and are available at different levels of aggregation. Such models are needed to address a broad class of environmental agents that vary over time and across geographical regions. The project focuses on developing new statistical methods for (1) estimating *temporal* associations between health outcomes and current and past environmental exposures when the underlying function is unknown and exposure has been measured with error; (2) estimating *spatial* associations between health outcomes and environmental exposures that properly take into account nonrandom sampling designs; and (3) conducting integrated analyses of *spatiotemporal* data on health and environmental exposures that take into account the sources of bias that arise from spatial and temporal aggregations. We shall apply some of the proposed statistical methods to data on air pollution, mortality, and temperature.

This research will provide a unified statistical framework for analyzing environmental epidemiologic data of practical importance. The work proposed here will contribute statistical methods to use in the field of environmental epidemiology; as we compile the data and apply the proposed methods to the various data sets, the results of the analyses will provide evidence on the relations among air pollution, temperature, and health effects.

** Project funded by HEI. The principal investigator is Jonathan M Samet.

†† Project funded by the National Institute of Environmental Health Sciences. The principal investigator is Francesca Dominici.

REFERENCES

Almon S. 1965. The distributed lag between capital appropriations and expenditures. Econometrica 33:178–196.

Burnett RT, Dewanji A, Dominici F, Goldberg MS, Cohen A, Krewski D. 2003. On the relationship between time-series studies, dynamic population studies, and estimating loss of life due to short-term exposure to environmental risks. Environ Health Perspect 111:1170–1174.

Burnett R, Krewski D. 1994. Air pollution effects of hospital admission rates: A random effects modeling approach. Can J Stat 22:441–458.

Burnett R, Ma R, Jerrett M, Goldberg MS, Cakmak S, Pope CA III, Krewski D. 2001. The spatial association between community air pollution and mortality: A new method of analyzing correlated geographic cohort data. Environ Health Perspect 109(Suppl 3):375–380.

Cakmak S, Burnett R, Krewski D. 1999. Methods for detecting and estimating population threshold concentrations for air pollution-related mortality with exposure measurement error. Risk Anal 19:487–496.

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

Carroll RJ, Ruppert D, Stefanski LA. 1995. Measurement Error in Nonlinear Models. Monographs on Statistics and Applied Probability Vol 63. CRC Press, Boca Raton FL.

Chambers JM, Hastie TJ, eds. 1992. Statistical Models in S. Wadsworth & Brooks, Pacific Grove CA.

Daniels MJ, 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 152:397–406.

Daniels MJ, Dominici F, Zeger S. 2004a. Underestimation of standard errors in multi site time series studies. Epidemiology 15[1]:57–62.

Daniels MJ, Dominici F, Zeger SL, Samet JM. 2004b. The National Morbidity, Mortality, and Air Pollution Study, Part III: PM_{10} Concentration–Response Curves and Threshold for the 20 Largest US Cities. Research Report 94. Health Effects Institute, Boston MA.

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. Control Clin Trials 7:177–188.

Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG Jr, Speizer FE. 1993. An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759.

Dominici F. 2002. Invited commentary. Air pollution and health: What can we learn from an hierarchical approach? Am J Epidemiol 155:11–15.

Dominici F, Daniels M, McDermott A, Zeger SL, Samet JM. 2003a. Revised analyses of selected time-series studies: Shape of the exposure–response relation and mortality displacement in the NMMAPS database. In: Revised Analyses of Time-Series Studies of Air Pollution and Health (pp 91–96). Special Report. Health Effects Institute, Boston MA.

Dominici F, Daniels M, Zeger SL, Samet JM. 2002a. 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. 2003b. Revised analyses of the National Morbidity, Mortality, and Air Pollution Study, Part II: 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(486):938–948.

Dominici F, McDermott A, Zeger SL, Samet JM. 2002b. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol 156: 193–203.

Dominici F, McDermott A, Zeger SL, Samet JM. 2003c. Airborne particulate matter and mortality: Timescale effects in four US cities. Am J Epidemiol 157:1055–1065.

Dominici F, McDermott A, Zeger SL, Samet JM. 2003d. National maps of the effects of particulate matter on mortality: Exploring geographical variation. Environ Health Perspect 111:39–43.

Dominici F, Samet JM, Zeger SL. 2000a. Combining evidence on air pollution and daily mortality from the 20 largest US cities: A hierarchical modeling strategy. J Royal Stat Soc Ser A 163:263–302.

Dominici F, Samet JM, Zeger SL. 2000b. Section 5: Combining evidence on air pollution and daily mortality from twenty largest US cities. In: The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues [pp 61–74]. Research Report 94. Health Effects Institute, Cambridge MA.

Dominici F, Sheppard L, Clyde M. 2003e. Health effects of air pollution: A statistical review. Int Stat Rev 71:243–276.

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.

Dominici F, Zeger SL, Samet JM. 2000c. A measurement error model for time-series studies of air pollution and mortality. Biostatistics 1:157–175.

Dominici F, Zeger SL, Samet JM. 2000d. Section 2: A measurement error model for time-series studies of air pollution and mortality. In: The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues [pp 29–41]. Research Report 94. Health Effects Institute, Cambridge MA.

Duan N. 1991. Stochastic microenvironment models for air pollution exposure. J Expos Anal Environ Epidemiol 1:235–257.

Gelman A, Carlin JB, Stern HS, Rubin DB. 1995. Bayesian Data Analysis. Chapman & Hall, London UK.

Gilks WR, Richardson S, Spiegelhalter DJ, eds. 1996. Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics. Chapman & Hall, London UK.

Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732.

Haran M, Carlin BP, Adgate JL, Ramachandran G, Waller LA, Gelfand AE. 2002. Hierarchical Bayes models for relating particulate matter exposure measures. In: Case Studies in Bayesian Statistics, Vol VI (Gatsonis C, Carriquiry A, Gelman A, Higdon D, Kaas R, Pauler D, Verdinelli I, eds). Springer-Verlag, New York NY.

Hastie TJ, Tibshirani RJ. 1990. Generalized Additive Models. Chapman & Hall, New York NY.

Health Effects Institute. 2003. Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. Health Effects Institute, Boston MA.

Janssen NAH, Hoek G, Brunekreef B, Harssema H, Mensink I, Zuidhof A. 1998. Personal sampling of particles in adults: Relation among personal, indoor, and outdoor air concentrations. Am J Epidemiol 147:537–544.

Janssen NAH, Hoek G, Harssema H, Brunekreef B. 1997. Childhood exposure to PM_{10}: Relation between personal, classroom, and outdoor concentrations. Occup Environ Med 54:888–894.

Katsouyanni K, Toulomi G, Spix C, Balducci F, Medina S, Rossi G, Wojtyniak B, Sunyer J, Bacharova L, Schouten JP, 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. BMJ 314:1658–1663.

Kelsall JE, Samet JM, Zeger SL, Xu J. 1997. Air pollution and mortality in Philadelphia, 1974–1988. Am J Epidemiol 146:750–762.

Kelsall JE, Zeger SL, Samet JM. 1999. Frequency domain log-linear models; air pollution and mortality. J Royal Stat Soc Ser C 48:331–344.

Klein M, Flanders WD, Tolbert PE. 2002. Variances may be underestimated using available software for generalized additive models. Abstract L1. Am J Epidemiol 155:S106.

Li Y, Roth HD. 1995. Daily mortality analysis by using different regression models in Philadelphia County, 1973–1990. Inhalation Toxicol 7:45–58.

Liang K-Y, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73:13–22.

Lindley DV, Smith AFM. 1972. Bayes estimates for the linear model (with discussion). J Royal Stat Soc Ser B 34:1–41.

Lioy PJ, Waldman JM, Buckley T, Butler J, Pietarinen C. 1990. The personal, indoor and outdoor concentrations of PM_{10} measured in an industrial community during the winter. Atmos Environ 24B:57–66.

Lipfert FW, Wyzga RE. 1995. Air pollution and mortality: Issues and uncertainties. J Air Waste Manage Assoc 45:949–966.

Lumley T, Sheppard L. 2003. Time series analyses of air pollution and health: Straining at gnats and swallowing camels? Epidemiology 14:13–14.

Mage T, Buckley TJ. 1995. The relationship between personal exposures and ambient concentrations of particulate matter. In: Proceedings of the 88th Annual Meeting of the Air & Waste Management Association, June 1995, Pittsburgh PA (pp 1–16).

McCullagh P, Nelder JA. 1989. Generalized Linear Models, 2nd edition. Chapman & Hall, New York NY.

Morris CN, Normand S-L. 1992. Hierarchical models for combining information and for meta-analysis. In: Bayesian Statistics 4 (Bernardo JM, Berger JO, Dawid AP, Smith AFM, eds), pp 321–344. Oxford University Press, Oxford, England.

National Research Council (US). 1998. Research Priorities for Airborne Particulate Matter: I. Immediate Priorities and a Long-Range Research Portfolio. National Academy Press, Washington DC.

National Research Council (US). 1999. Research Priorities for Airborne Particulate Matter: II. Evaluating Research Progress and Updating the Portfolio. National Academy Press, Washington DC.

National Research Council (US). 2001. Research Priorities for Airborne Particulate Matter: III. Early Research Progress. National Academy Press, Washington DC.

Ott W, Wallace L, Mage D. 2000. Predicting particulate (PM_{10}) personal exposure distributions using a random component superposition model. J Air Waste Manage Assoc 50:1390–1406.

Ozkaynak H, Xue J, Spengler J, Wallace L, Pellizzari E, Jenkins P. 1996. Personal exposure to airborne particles and metals: Results from the Particle TEAM study in Riverside, California. J Expos Anal Environ Epidemiol 6:57–78.

Pope CA III. 2000. Invited commentary: Particulate matter-mortality exposure-response relations and threshold. Am J Epidemiol 152:407–412.

Pope CA III, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW Jr. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir Crit Care Med 151:669–674.

Ramsay TO, Burnett RT, Krewski D. 2003. The effect of concurvity in generalized additive models linking mortality and ambient particulate matter. Epidemiology 14:18–23.

Rodes CE, Lawless PA, Evans GF, Sheldon LS, Williams RW, Vette AF, Creason JP, Walsh D. 2001. The relationships between personal PM exposures for elderly populations and indoor and outdoor concentrations for three retirement center scenarios. J Expos Anal Environ Epidemiol 11:103–115.

Roemer W, Hoek G, Brunekreef B, Haluszka J, Kalandidi A, Pekkanen J. 1998. Daily variations in air pollution and respiratory health in a multicentre study: the PEACE project. Pollution Effects on Asthmatic Children in Europe. Eur Respir J 12:1354–1361.

Samet JM, Dominici F, Curriero F, Coursac I, Zeger SL. 2000a. Fine particulate air pollution and mortality in 20 US cities, 1987–1994 (with discussion). N Engl J Med 343:1742–1749.

Samet JM, Dominici F, McDermott A, Zeger SL. 2003. New problems for an old design: Time series analyses of air pollution and health. Epidemiology 14:11–12.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000b. Section 1: Exposure measurement error in time-series studies of air pollution. In: The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues [pp 15–27]. Research Report 94. Health Effects Institute, Cambridge MA.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000c. Section 3: Mortality displacement–resistant estimates of air pollution effects on mortality. In: The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues [pp 43–49]. Research Report 94. Health Effects Institute, Cambridge MA.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000d. The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodological Issues (pp 5–14). Research Report 94. Health Effects Institute, Cambridge MA.

Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, Dockery DW, Schwartz J, Zanobetti A. 2000e. The National Morbidity, Mortality, and Air Pollution Study, Part II: Morbidity and Mortality from Air Pollution in the United States (pp 5–6). Research Report 94. Health Effects Institute, Cambridge MA.

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

Schwartz J. 2000b. Harvesting and long term exposure effects in the relation between air pollution and mortality. Am J Epidemiol 151:440–448.

Schwartz J. 2001. Is there harvesting in the association of airborne particles with daily deaths and hospital admissions? Epidemiology 12:55–61.

Schwartz J, Zanobetti A. 2000. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. Epidemiology 6:666–672.

Sheppard L, Damian D. 2000. Estimating short-term PM effects accounting for surrogate exposure measurements from ambient monitors. Environmetrics 11:675–687.

Smith RL, Spitzner D, Kim Y, Fuentes M. 2000. Threshold dependence of mortality effects for fine and coarse particles in Phoenix, Arizona. J Air Waste Manage Assoc 50:1367–1379.

Tanner MA. 1991. Tools for Statistical Inference: Observed Data and Data Augmentation Methods, Vol 67 of Lecture Notes in Statistics. Springer-Verlag, New York NY.

Thomas A, Spiegelhalter DJ, Gilks WR. 1992. BUGS: A program to perform Bayesian inference using Gibbs sampling. In: Bayesian Statistics 4 (Bernardo JM, Berger JO, Dawid AP, Smith AFM, eds), pp 837–842. Oxford University Press, Oxford, England.

Thomas D, Stram D, Dwyer J. 1993. Exposure measurement error: Influence on exposure-disease relationships and methods of corrections. Annu Rev Public Health 14:69–93.

Tierney L. 1994. Markov chains for exploring posterior distributions (with discussion). Ann Stat 22:1701–1762.

Toulomi G, Katsouyanni K, Zmirou D, Schwartz J, Spix C, Ponce de Leon A, Tobias A, Quennel P, Rabczenko D, Bacharova L, Bisanti L, Vonk JM, Ponka A. 1997. Short-term effects of ambient oxidant exposure on mortality: A combined analysis within the APHEA project. Am J Epidemiol 146:177–185.

Wilson WE, Mage DT, Grant LD. 2000. Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: Why and how. J Air Waste Manage Assoc 50:1167–1183.

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

Zanobetti A, Schwartz J, Samoli E, Gryparis A, Touloumi G, Atkinson R, Le Tertre A, Bobros J, Celko M, Goren A, Forsberg B, Michelozzi P, Rabczenko D, Aranguez Ruiz E, Katsouyanni K. 2002. The temporal pattern of mortality responses to air pollution: A multicity assessment of mortality displacement. Epidemiology 13:87–93.

Zanobetti A, Wand MP, Schwartz J, Ryan LM. 2000b. Generalized additive distributed lag models: Quantifying mortality displacement. Biostatistics 1:279–292.

Zeger SL, Dominici F, Samet J. 1999. Harvesting-resistant estimates of pollution effects on mortality. Epidemiology 10:171–175.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, Cohen A. 2000. Exposure measurement error in time-series studies of air pollution: Concepts and consequences. Environ Health Perspect 108:419–426.

APPENDIX A. Selected Abstracts

This appendix presents abstracts of selected articles and reports that either formed the basis for this work or report the findings of research conducted by Francesca Dominici and her colleagues under her contract as the recipient of the Walter A Rosenblith New Investigator Award. (All references cited within an abstract refer to the list of references in the original publications and not to the list included with this Investigator's Report.)

UNDERESTIMATION OF STANDARD ERRORS IN MULTI SITE TIME SERIES STUDIES (*Michael J Daniels, Francesca Dominici, and Scott L Zeger*; reprinted from Daniels et al 2004a)

Multi-site time series studies of the association of air pollution with mortality and morbidity have figured prominently in the literature as comprehensive approaches for estimating short-term effects of air pollution on health. Hierarchical models are generally used to combine site-specific information and to estimate pooled air pollution effects while taking into account both within-site statistical uncertainty and across-site heterogeneity.

Within a site, characteristics of time series data of air pollution and health (small pollution effects, missing data, and highly correlated predictors) make modeling of all sources of uncertainty challenging. One potential consequence is underestimation of the statistical variance of the site-specific effects to be combined.

In this paper, we investigate the impact of variance underestimation on the pooled relative rate estimate. We focused on two-stage normal-normal hierarchical models and on underestimation of the statistical variance at the first stage. By mathematical considerations and simulation studies, we found that variance underestimation did not affect the pooled estimate substantially. However, the pooled estimate was somewhat sensitive to variance underestimation when the number of sites was small and underestimation was severe. These simulation results are applicable to any two-stage normal-normal hierarchical model for combining information of site-specific results (including meta-analyses), and they can be easily extended to more general hierarchical formulations.

We also examined the impact of variance underestimation on the national average relative rate estimate from the National Morbidity, Mortality and Air Pollution Study. We found that variance underestimation as large as 40% had little effect on the national average.

IMPROVED SEMI-PARAMETRIC TIME SERIES MODELS OF AIR POLLUTION AND MORTALITY (*Francesca Dominici, Aidan McDermott, and Trevor J Hastie*; reprinted from Dominici et al 2004a)

In 2002, methodological issues around time series analyses of air pollution and health attracted the attention of the scientific community, policy makers, the press, and the diverse stakeholders concerned with air pollution. As the Environmental Protection Agency (EPA) was finalizing its most recent review of epidemiological evidence on particulate matter air pollution (PM), statisticians and epidemiologists found that the S-Plus implementation of Generalized Additive Models (GAM) can overestimate effects of air pollution and understate statistical uncertainty in time series studies of air pollution and health. This discovery delayed the completion of the PM Criteria Document prepared as part of the review of the U.S. National Ambient Air Quality Standard (NAAQS), as the time-series findings were a critical component of the evidence. In addition, it raised concerns about the adequacy of current model formulations and their software implementations.

In this paper we provide improvements in semi-parametric regression directly relevant to risk estimation in time series studies of air pollution. First, we introduce a closed form estimate of the asymptotically exact covariance matrix of the linear component of a GAM. To ease the implementation of these calculations, we develop the S package GAM.EXACT, an extended version of *gam*. Use of GAM.EXACT allows a more robust assessment of the statistical uncertainty of the estimated pollution coefficients. Second, we develop a bandwidth selection method to reduce confounding bias in the pollution-mortality relationship due to unmeasured time-varying factors such as season and influenza epidemics. Our method selects the number of degrees of freedom in the smooth part of the model that minimizes the mean squared error of the air pollution coefficient. Third, we introduce a conceptual framework to fully explore the sensitivity of the air pollution risk estimates to model choice. We apply our methods to data of the National Mortality Morbidity Air Pollution Study (NMMAPS), which includes time series data from the 90 largest US cities for the period 1987–1994.

HIERARCHICAL BIVARIATE TIME SERIES MODELS: A COMBINED ANALYSIS OF THE EFFECTS OF PARTICULATE MATTER ON MORBIDITY AND MORTALITY (*Francesca Dominici, Antonella Zanobetti, Scott L Zeger, Joel Schwartz, and Jonathan M Samet*; reprinted from Dominici et al 2004b)

In this paper we develop a hierarchical bivariate time series model to characterize the relationship between particulate matter less than 10 microns in aerodynamic diameter (PM_{10}) and both mortality and hospital admissions for cardiovascular diseases. The model is applied to time series data on mortality and morbidity for 10 metropolitan areas in the United States from 1986 to 1993. We postulate that these time series should be related through a shared relationship with PM_{10}.

At the first stage of the hierarchy, we fit two seemingly unrelated Poisson regression models to produce city-specific estimates of the log relative rates of mortality and morbidity associated with exposure to PM_{10} within each location. The sample covariance matrix of the estimated log relative rates is obtained using a novel generalized estimating equation approach that takes into account the correlation between the mortality and morbidity time series. At the second stage, we combine information across locations to estimate overall log relative rates of mortality and morbidity and variation of the rates across cities.

Using the combined information across the 10 locations we find that a 10 $\mu g/m^3$ increase in average PM_{10} at the current day and previous day is associated with a 0.26% increase in mortality (95% posterior interval −0.37, 0.65), and a 0.71% increase in hospital admissions (95% posterior interval 0.35, 0.99). The log relative rates of mortality and morbidity have a similar degree of heterogeneity across cities: the posterior means of the between-city standard deviations of the mortality and morbidity air pollution effects are 0.42 (95% interval 0.05, 1.18), and 0.31 (95% interval 0.10, 0.89), respectively. The city-specific log relative rates of mortality and morbidity are estimated to have very low correlation, but the uncertainty in the correlation is very substantial (posterior mean = 0.20, 95% interval −0.89, 0.98).

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

ON THE RELATIONSHIP BETWEEN TIME-SERIES STUDIES, DYNAMIC POPULATION STUDIES, AND ESTIMATING LOSS OF LIFE DUE TO SHORT-TERM EXPOSURE TO ENVIRONMENTAL RISKS
(*Richard T Burnett, Anup Dewanji, Francesca Dominici, Mark S Goldberg, Aaron Cohen, and Daniel Krewski*; reprinted from Burnett et al 2003)

There is a growing concern that short-term exposure to combustion-related air pollution is associated with increased risk of death. This finding is based largely on time-series studies that estimate associations between daily variations in ambient air pollution concentrations and in the number of nonaccidental deaths within a community. Because these results are not based on cohort or dynamic population designs, where individuals are followed in time, it has been suggested that estimates of effect from these time-series studies cannot be used to determine the amount of life lost because of short-term exposures. We show that results from time-series studies are equivalent to estimates obtained from a dynamic population when each individual's survival experience can be summarized as the daily number of deaths. This occurs when the following conditions are satisfied: a) the environmental covariates vary in time and not between individuals; b) on any given day, the probability of death is small; c) on any given day and after adjusting for known risk factors for mortality such age, sex, smoking habits, and environmental exposures, each subject of the at-risk population has the same probability of death; d) environmental covariates have a common effect on mortality of all members of at-risk population; and e) the averages of individual risk factors, such as smoking habits, over the at-risk population vary smoothly with time. Under these conditions, the association between temporal variation in the environmental covariates and the survival experience of members of the dynamic population can be estimated by regressing the daily number of deaths on the daily value of the environmental covariates, as is done in time-series mortality studies. Issues in extrapolating risk estimates based on time-series studies in one population to estimate the amount of life lost in another population are also discussed.

SHAPE OF THE EXPOSURE–RESPONSE RELATION AND MORTALITY DISPLACEMENT IN THE NMMAPS DATABASE (*Francesca Dominici, Michael Daniels, Aidan McDermott, Scott L Zeger, and Jonathan M Samet*; reprinted from Dominici et al 2003a)

This section summarizes the revised results of three analyses of the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) data for particulate matter (PM) less than 10 µm in aerodynamic diameter (PM_{10}) and mortality: (1) estimation of shape of the exposure–response relation and location of any threshold among the 20 largest US cities (1987–1994); (2) estimation of regional and national average exposure–response curves among the 88 largest US cities (1987–1994); and (3) frequency and time-domain log-linear regression analyses to assess the extent of mortality displacement in Philadelphia (1974–1987). These analyses did not show any substantial differences from previous analyses.

MORTALITY AMONG RESIDENTS OF 90 CITIES (*Francesca Dominici, Aidan McDermott, Michael Daniels, Scott L Zeger, and Jonathan M Samet*; reprinted from Dominici et al 2003b)

This report presents findings from updated analyses of data from 90 US cities assembled for the National Morbidity, Mortality and Air Pollution Study (NMMAPS). The data were analyzed with a generalized additive model (GAM) using the *gam* function in S-Plus (with default convergence criteria previously used and with more stringent criteria) and with a generalized linear model (GLM) with natural cubic splines. With the original method, the estimated effect of PM_{10} (particulate matter 10 µm in mass median aerodynamic diameter) on total mortality from nonexternal causes was a 0.41% increase per 10 µg/m^3 increases in PM_{10}; with the more stringent criteria, the estimate was 0.27%; and with GLM, the effect was 0.21%. The effect of PM_{10} on respiratory and cardiovascular mortality combined was greater, but the pattern across models was similar. The findings of the updated analysis with regard to spatial heterogeneity across the 90 cities were unchanged from the original analyses.

AIRBORNE PARTICULATE MATTER AND MORTALITY: TIMESCALE EFFECTS IN FOUR US CITIES (*Francesca Dominici, Aidan McDermott, Scott L Zeger, and Jonathan M Samet*; reprinted from Dominici et al 2003c)

While time-series studies have consistently provided evidence for an effect of particulate air pollution on mortality, uncertainty remains as to the extent of the life-shortening implied by those associations. In this paper, the authors estimate the association between air pollution and mortality using different timescales of variation in the air pollution time series to gain further insight into this question. The authors' method is based on a Fourier decomposition of air pollution time series into a set of independent exposure variables, each representing a different timescale. The authors then use this set of variables as predictors in a Poisson regression model to estimate a separate relative rate of mortality for each exposure timescale. The method is applied to a database containing information on daily mortality, particulate air pollution, and weather in four US cities (Pittsburgh, Pennsylvania; Minneapolis, Minnesota; Seattle, Washington; and Chicago, Illinois) from the period 1987–1994. The authors found larger relative rates of mortality associated with particulate air pollution at longer timescale variations (14 days–2 months) than at shorter timescales (1–4 days). These analyses provide additional evidence that associations between particle indexes and mortality do not imply only an advance in the timing of death by a few days for frail individuals.

NATIONAL MAPS OF THE EFFECTS OF PARTICULATE MATTER ON MORTALITY: EXPLORING GEOGRAPHICAL VARIATION (*Francesca Dominici, Aidan McDermott, Scott L Zeger, and Jonathan M Samet*; reprinted from Dominici et al 2003d)

In this paper, we present national maps of relative rates of mortality associated with short-term exposure to particulate matter < 10 µm in aerodynamic diameter (PM_{10}). We report results for 88 of the largest metropolitan areas in the United States from 1987 to 1994 for all-cause mortality, combined cardiovascular and respiratory deaths, and other causes of mortality. Maximum likelihood estimates of the relative rate of mortality associated with PM_{10} and the degree of statistical uncertainty were obtained for each of the 88 cities by fitting a separate log-linear regression of the daily mortality rate on air pollution level and potential confounders. We obtained Bayesian estimates of the relative rates by fitting a hierarchical model that takes into account spatial correlation among the true city-specific relative rates. We found that daily variations of PM_{10} are positively associated with daily variations of mortality. In particular, the relative rate estimates of cardiovascular and respiratory mortality associated with PM_{10} are larger on average than the relative rate estimates of all-cause and other-cause mortality. The estimated increase in the relative rate of death from cardiovascular and respiratory mortality, all-cause mortality, and other-cause mortality were 0.31% (95% posterior interval, 0.15–0.5), 0.22% (95% posterior interval, 0.1–0.38), and 0.13% (95% posterior interval, −0.05 to 0.29), respectively. Bayesian estimates of the city-specific relative rates ranged from 0.23% to 0.35% for cardiovascular and respiratory mortality, from 0.18% to 0.27% for all causes, and from 0.10% to 0.20% for other causes of mortality. The spatial characterization of effects across cities offers the potential to identify factors that could influence the effect of PM_{10} on health, including particle characteristics, offering insights into mechanisms by which PM_{10} causes adverse health effects.

HEALTH EFFECTS OF AIR POLLUTION: A STATISTICAL REVIEW (*Francesca Dominici, Lianne Sheppard, and Merlise Clyde*; reprinted from Dominici et al 2003e)

We critically review and compare epidemiological designs and statistical approaches to estimate associations between air pollution and health. More specifically, we aim to address the following questions:

1. Which epidemiological designs and statistical methods are available to estimate associations between air pollution and health?
2. What are the recent methodological advances in the estimation of the health effects of air pollution in time series studies?
3. What are the main methodological challenges and future research opportunities relevant to regulatory policy?

In question 1, we identify strengths and limitations of time series, cohort, case-crossover and panel sampling designs. In question 2, we focus on time series studies and we review statistical methods for: 1) combining information across multiple locations to estimate overall air pollution effects; 2) estimating the health effects of air pollution taking into account of model uncertainties; 3) investigating the consequences of exposure measurement error in the estimation of the health effects of air pollution; and 4) estimating air pollution-health exposure-response curves. Here, we also discuss the extent to which these statistical contributions have addressed key substantive questions. In question 3, within a set of policy-relevant-questions, we identify research opportunities and point out current data limitations.

AIR POLLUTION AND MORTALITY: ESTIMATING REGIONAL AND NATIONAL DOSE–RESPONSE RELATIONSHIPS (*Francesca Dominici, Michael Daniels, Scott L Zeger, and Jonathan M Samet*; reprinted from Dominici et al 2002a)

We analyzed a national data base of air pollution and mortality for the 88 largest US cities for the period 1987–1994, to estimate relative rates of mortality associated with airborne particulate matter less than 10 microns (PM_{10}), and the form of the relationship between PM_{10} concentration and mortality. To estimate city-specific relative rates of mortality associated with PM_{10}, we built log-linear models, which included nonparametric adjustments for weather variables and longer term trends. To estimate PM_{10}-mortality dose–response curves, we modeled the logarithm of the expected value of daily mortality as a function of PM_{10} using natural cubic splines with unknown numbers and locations of knots. We also developed spatial models to investigate the heterogeneity of relative mortality rates and of the shapes of PM_{10}-mortality dose–response curves across cities and geographical regions. To determine whether variability in effect estimates can be explained by city-specific factors, we explored the dependence of relative mortality rates on mean pollution levels, demographic variables, reliability of the pollution data, and specific constituents of particulate matter. We implemented estimation with simulation-based methods, including data-augmentation to impute the missing data of the city-specific covariates, and the reversible jump Monte Carlo Markov chain (RJMCMC) to sample from the posterior distribution of the parameters in the hierarchical spline model. We found that previous day PM_{10} concentrations were positively associated with total mortality in the majority of the locations, with a 0.5% increment for a 10 $\mu g/m^3$ increase in PM_{10}. The effect was strongest in the Northeast region, where the increase in the death rate was twice as high as the average for the other cities. Overall, we found that the pooled concentration-response relationship for the nation was linear.

ON THE USE OF GENERALIZED ADDITIVE MODELS IN TIME-SERIES STUDIES OF AIR POLLUTION AND HEALTH (*Francesca Dominici, Aidan McDermott, Scott L Zeger, and Jonathan M Samet*; reprinted from Dominici et al 2002b)

The widely used generalized additive models (GAM) method is a flexible and effective technique for conducting nonlinear regression analysis in time-series studies of the health effects of air pollution. When the data to which the GAM are being applied have two characteristics—1) the estimated regression coefficients are small and 2) there exist confounding factors that are modeled using at least two nonparametric smooth functions—the default settings in the *gam* function of the S-Plus software package (version 3.4) do not assure convergence of its iterative estimation procedure and can provide biased estimates of regression coefficients and standard errors. This phenomenon has occurred in time-series analyses of contemporary data on air pollution and mortality. To evaluate the impact of default implementation of the *gam* software on published analyses, the authors reanalyzed data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) using three different methods: 1) Poisson regression with parametric nonlinear adjustments for confounding factors; 2) GAM with default convergence parameters; and 3) GAM with more stringent convergence parameters than the default settings. The authors found that pooled NMMAPS estimates were very similar under the first and third methods but were biased upward under the second method.

ESTIMATING PARTICULATE MATTER-MORTALITY DOSE-RESPONSE CURVES AND THRESHOLD LEVELS: AN ANALYSIS OF DAILY TIME-SERIES FOR THE 20 LARGEST US CITIES (*Michael J Daniels, Francesca Dominici, Jonathan M Samet, and Scott L Zeger*; reprinted from Daniels et al 2000)

Numerous studies have shown a positive association between daily mortality and particulate air pollution, even at concentrations below regulatory limits. These findings have motivated interest in the shape of the exposure-response relation. The authors have developed flexible modeling strategies for time-series data that include spline and threshold exposure-response models; they apply these models to daily time-series data for the 20 largest US cities for 1987–1994, using the concentration of particulate matter < 10 µm in aerodynamic diameter (PM_{10}) as the exposure measure. The spline model showed a linear relation without indication of threshold for PM_{10} and relative risk of death for all causes and cardiorespiratory causes; by contrast, for other causes, the risk did not increase until approximately 50 µg/m^3 PM_{10}. For all-cause mortality, a linear model without threshold was preferred to the threshold model and to the spline model, using the Akaike information criterion (AIC). The findings were similar for cardiovascular and respiratory deaths combined. By contrast, for causes other than cardiovascular and respiratory, a threshold model was more competitive with a threshold value estimated at 65 µg/m^3. These findings indicate that linear models without a threshold are appropriate for assessing the effect of particulate air pollution on daily mortality even at current levels.

SECTION 5: COMBINING EVIDENCE ON AIR POLLUTION AND DAILY MORTALITY FROM TWENTY LARGEST US CITIES (*Francesca Dominici, Jonathan M Samet, and Scott L Zeger*; reprinted from Dominici et al 2000b)

Reports over the last decade of association between levels of particles in outdoor air and daily mortality counts have raised concern that air pollution shortens life, even at concentrations within current regulatory limits. Criticisms of these reports have focused on the statistical techniques used to estimate the pollution-mortality relationship and the inconsistency in findings among cities. We have developed analytic methods that address these concerns and combine evidence from multiple locations in order to gain a unified analysis of the data.

Section 5 introduces hierarchical regression models for combining estimates of the pollution-mortality relationship across cities and presents log-linear regression analyses of daily time-series data from the 20 largest US cities as an example. We illustrate this method focusing on health effects of particulate matter less than 10 µm in aerodynamic diameter (PM_{10}) and considering univariate and bivariate analyses with PM_{10} and O_3. In the first stage of the hierarchical model, we estimate the relative mortality rate associated with PM_{10} and O_3 for each of the 20 cities using semiparametric log-linear models. The second stage of the model describes between-city variation in the true relative rates as a function of selected city-specific covariates. We also fit 2 variations of a spatial model with the goal of exploring the spatial correlation of the pollutant-specific coefficients among cities. Finally, to explore the results of considering the 2 pollutants jointly, we fit and compared univariate and bivariate models. All posterior distributions from stage 2 are estimated using Markov chain Monte Carlo (MCMC) techniques. Results appear to be largely insensitive to the specific choice of vague but proper prior distribution. The models and estimation methods are general and can be used for any number of locations and pollutant measurements and have potential application to other environmental agents.

COMBINING EVIDENCE ON AIR POLLUTION AND DAILY MORTALITY FROM THE 20 LARGEST US CITIES: A HIERARCHICAL MODELLING STRATEGY (*Francesca Dominici, Jonathan M Samet, and Scott L Zeger*; reprinted from Dominici et al 2000a)

[Read before *The Royal Statistical Society* on Wednesday January 12th, 2000, the President, *Professor DA Lievesley*, in the Chair]

Reports over the last decade of association between levels of particles in outdoor air and daily mortality counts have raised concern that air pollution shortens life, even at concentrations within current regulatory limits. Criticisms of these reports have focused on the statistical techniques that are used to estimate the pollution–mortality relationship and the inconsistency in findings between cities. We have developed analytical methods that address these concerns and combine evidence from multiple locations to gain a unified analysis of the data. The paper presents log-linear regression analyses of daily time series data from the largest 20 US cities and introduces hierarchical regression models for combining estimates of the pollution–mortality relationship across cities. We illustrate this method by focusing on mortality effects of PM_{10} (particulate matter less than 10 μm in aerodynamic diameter) and by performing univariate and bivariate analyses with PM_{10} and ozone (O_3) level. In the first stage of the hierarchical model, we estimate the relative mortality rate associated with PM_{10} for each of the 20 cities by using semiparametric log-linear models. The second stage of the model describes between-city variation in the true relative rates as a function of selected city-specific covariates. We also fit two variations of a spatial model with the goal of exploring the spatial correlation of the pollutant-specific coefficients among cities. Finally, to explore the results of considering the two pollutants jointly, we fit and compare univariate and bivariate models. All posterior distributions from the second stage are estimated by using Markov chain Monte Carlo techniques. In univariate analyses using concurrent day pollution values to predict mortality, we find that an increase of 10 μg m^{-3} in PM_{10} on average in the USA is associated with a 0.48% increase in mortality (95% interval: 0.05, 0.92). With adjustment for the O_3 level the PM_{10}-coefficient is slightly higher. The results are largely insensitive to the specific choice of vague but proper prior distribution. The models and estimation methods are general and can be used for any number of locations and pollutant measurements and have potential applications to other environmental agents.

SECTION 2: A MEASUREMENT ERROR MODEL FOR TIME-SERIES STUDIES OF AIR POLLUTION AND MORTALITY (*Francesca Dominici, Scott L Zeger, and Jonathan M Samet*; reprinted from Dominici et al 2000d)

One barrier to interpreting the observational evidence concerning the adverse health effects of air pollution for public policy purposes is the measurement error inherent in estimates of exposure based on ambient pollutant monitors. Exposure assessment studies have shown that data from monitors at central sites may not adequately represent personal exposure. Thus, the exposure error resulting from using centrally measured data as a surrogate for personal exposure can potentially lead to bias in estimates of the health effects of air pollution.

This section of the Investigators' Report presents a multistage Poisson regression model for evaluating the effects of exposure measurement error on estimates of effects of ambient particulate matter (PM) on mortality in time-series studies. To implement the model, we have used 5 validation data sets on personal exposure to PM less than 10 μm in aerodynamic diameter (PM_{10}). Our goal is to combine data on the associations between ambient concentrations of PM and mortality for a specific location, with the validation data on the association between ambient and personal concentrations of PM at the locations where data have been collected. We use these data in a model to estimate the relative risk of mortality associated with estimated personal exposure concentrations and compare this estimate with the risk of mortality estimated with measurements of ambient concentration alone. We apply this method to data comprising daily mortality counts, ambient concentrations of PM_{10} measured at a central site, and temperature for Baltimore, Maryland, from 1987 to 1994. We have selected our home city of Baltimore to illustrate the method; the measurement error correction model is general and can be applied to other appropriate locations.

Our approach uses a combination of (1) a generalized additive model with log link and Poisson error for the mortality–personal exposure association, (2) a multistage linear model to estimate the variability across the 5 validation data sets in the personal–ambient exposure association, and (3) data augmentation to address the uncertainty resulting from the missing personal exposure time series in Baltimore. In the Poisson regression model, we account for smooth seasonal and annual trends in mortality using smoothing splines. Taking into account the heterogeneity across locations in the personal–ambient exposure relationship, we quantify the degree to which the exposure measurement error biases the results toward the null hypothesis of no effect, and estimate the loss of precision in the estimated health effects due to indirectly estimating personal exposures from ambient measurements.

A MEASUREMENT ERROR MODEL FOR TIME-SERIES STUDIES OF AIR POLLUTION AND MORTALITY

(*Francesca Dominici, Scott L Zeger, and Jonathan M Samet*; reprinted from Dominici et al 2000c)

One barrier to interpreting the observational evidence concerning the adverse health effects of air pollution for public policy purposes is the measurement error inherent in estimates of exposure based on ambient pollutant monitors. Exposure assessment studies have shown that data from monitors at central sites may not adequately represent personal exposure. Thus, the exposure error resulting from using centrally measured data as a surrogate for personal exposure can potentially lead to a bias in estimates of the health effects of air pollution.

This paper develops a multi-stage Poisson regression model for evaluating the effects of exposure measurement error on estimates of effects of particulate air pollution on mortality in time-series studies. To implement the model, we have used five validation data sets on personal exposure to PM_{10}. Our goal is to combine data on the associations between ambient concentrations of particulate matter and mortality for a specific location, with the validation data on the association between ambient and personal concentrations of particulate matter at the locations where data have been collected. We use these data in a model to estimate the relative risk of mortality associated with estimated personal-exposure concentrations and make a comparison with the risk of mortality estimated with measurements of ambient concentration alone. We apply this method to data comprising daily mortality counts, ambient concentrations of PM_{10} measured at a central site, and temperature for Baltimore, Maryland from 1987 to 1994. We have selected our home city of Baltimore to illustrate the method; the measurement error correction model is general and can be applied to other appropriate locations.

Our approach uses a combination of: (1) a generalized additive model with log link and Poisson error for the mortality–personal-exposure association; (2) a multi-stage linear model to estimate the variability across the five validation data sets in the personal–ambient-exposure association; (3) data augmentation methods to address the uncertainty resulting from the missing personal exposure time series in Baltimore. In the Poisson regression model, we account for smooth seasonal and annual trends in mortality using smoothing splines. Taking into account the heterogeneity across locations in the personal–ambient-exposure relationship, we quantify the degree to which the exposure measurement error biases the results toward the null hypothesis of no effect, and estimate the loss of precision in the estimated health effects due to indirectly estimating personal exposures from ambient measurements.

FINE PARTICULATE AIR POLLUTION AND MORTALITY IN 20 U.S. CITIES, 1987–1994
(*Jonathan M Samet, Francesca Dominici, Frank C Curriero, Ivan Coursac, and Scott L Zeger*; reprinted from Samet et al 2000a)

Background. Air pollution in cities has been linked to increased rates of mortality and morbidity in developed and developing countries. Although these findings have helped lead to a tightening of air-quality standards, their validity with respect to public health has been questioned.

Methods. We assessed the effects of five major outdoor-air pollutants on daily mortality rates in 20 of the largest cities and metropolitan areas in the United States from 1987 to 1994. The pollutants were particulate matter that is less than 10 µm in aerodynamic diameter (PM_{10}), ozone, carbon monoxide, sulfur dioxide, and nitrogen dioxide. We used a two-stage analytic approach that pooled data from multiple locations.

Results. After taking into account potential confounding by other pollutants, we found consistent evidence that the level of PM_{10} is associated with the rate of death from all causes and from cardiovascular and respiratory illnesses. The estimated increase in the relative rate of death from all causes was 0.51 percent (95 percent posterior interval, 0.07 to 0.93 percent) for each increase in the PM_{10} level of 10 µg per cubic meter. The estimated increase in the relative rate of death from cardiovascular and respiratory causes was 0.68 percent (95 percent posterior interval, 0.20 to 1.16 percent) for each increase in the PM_{10} level of 10 µg per cubic meter. There was weaker evidence that increases in ozone levels increased the relative rates of death during the summer, when ozone levels are highest, but not during the winter. Levels of the other pollutants were not significantly related to the mortality rate.

Conclusions. There is consistent evidence that the levels of fine particulate matter in the air are associated with the risk of death from all causes and from cardiovascular and respiratory illnesses. These findings strengthen the rationale for controlling the levels of respirable particles in outdoor air.

THE NATIONAL MORBIDITY, MORTALITY, AND AIR POLLUTION STUDY, PART I: METHODS AND METHODOLOGIC ISSUES (*Jonathan M Samet, Francesca Dominici, Scott L Zeger, Joel Schwartz, and Douglas W Dockery*; reprinted from Samet et al 2000d)

OVERVIEW OF STUDY DESIGN AND CONCLUSIONS

The NMMAPS Project

The National Morbidity, Mortality, and Air Pollution Study (NMMAPS) comprises a comprehensive set of analyses of air pollution, mortality, and morbidity in a national sampling frame based on the monitoring information maintained in the US Environmental Protection Agency's (EPA's) Aerometric Information Retrieval System (AIRS). The project is a collaboration between investigators at The Johns Hopkins School of Public Health (Drs Samet, Zeger, and Dominici) and Harvard School of Public Health (Drs Dockery and Schwartz). The project's overall objectives lie in the complementary domains of methods development and methods application. This report, the first of two parts, details the methodologic components of NMMAPS. Part II provides the substantive findings on air pollution, mortality, and morbidity (Samet et al 2000).

The objectives for developing specific methodologic components for NMMAPS are fivefold.

1. To develop semiautomated or automated approaches for database construction using databases of the EPA, the National Center for Health Statistics (NCHS), the Health Care Financing Administration (HCFA), the Census Bureau, and the National Weather Service;
2. To develop and apply statistical methods for regression analyses of the multisite data, and to develop spatial time-series methods to estimate spatial maps of the relative rates of mortality associated with air pollution, while accounting, as necessary, for the spatial and temporal correlations in the mortality data;
3. To develop and apply methods that adjust for smooth trends and seasonality on mortality caused by changing demographics and health behaviors, influenza epidemics, and other unidentified factors;
4. To examine the consequences of measurement error in the exposure variables for assessing pollutant-mortality associations; and
5. To examine the degree to which pollution-related mortality reduces years of life (mortality displacement).

The objectives for application of methods developed for NMMAPS are threefold.

1. To assess the relation between air pollution and mortality in the largest US cities monitored for PM_{10} from 1987 forward;
2. To assess the relation between air pollution and morbidity in selected US cities monitored for PM_{10} from 1987 forward; and
3. To conduct paired analyses of morbidity and mortality in the same locations.

The design for NMMAPS builds on prior work supported by the Health Effects Institute in the Particle Epidemiology Evaluation Project (PEEP) (Samet et al 1995, 1997). This project was initiated in 1994 with the objectives of validating the data and replicating the findings in several of the time-series studies of air pollution and mortality reported during the 1990s. In a second phase, PEEP addressed several methodologic issues. These included selecting the approach for controlling for potential confounding by weather (Samet et al 1998) and determining the sensitivity of findings to model-building strategies (Kelsall et al 1997; Samet et al 1997).

The present project, NMMAPS, evolved from PEEP. The objectives encompassed methodologic issues that were persistent sources of uncertainty in interpreting the epidemiologic evidence: mortality displacement and exposure measurement error. The plan for multicity analyses was prompted by questioning the rationale for the study locations previously selected and by the prospect of setting this concern aside with analyses conducted using a defined sampling frame. Additionally, advances in hardware and software made this type of analysis feasible. The NMMAPS project was initiated at the end of 1996, as PEEP was ending.

THE NATIONAL MORBIDITY, MORTALITY, AND AIR POLLUTION STUDY PART II: MORBIDITY AND MORTALITY FROM AIR POLLUTION IN THE UNITED STATES (*Jonathan M Samet, Scott L Zeger, Francesca Dominici, Frank Curriero, Ivan Coursac, Douglas W Dockery, Joel Schwartz, and Antonella Zanobetti*; reprinted from Samet et al 2000e)

OVERVIEW

Project Objectives

The National Morbidity, Mortality, and Air Pollution Study (NMMAPS) comprises analyses of air pollution and mortality and morbidity set in a national sampling frame created from the Aerometric Information Retrieval System (AIRS), the monitoring database of the US Environmental Protection Agency. The project is a collaboration between investigators at The Johns Hopkins University School of Public Health (Drs Samet, Zeger, and Dominici) and Harvard School of Public Health (Drs Dockery, Schwartz, and Zanobetti). The project's overall objectives lie in the complementary domains of methods development and methods application.

This Report, Part II of NMMAPS, presents the findings on air pollution and morbidity and mortality in detail. For daily mortality, we have analyzed data for the 20 and 90 largest cities in the United States. (Throughout this Report, we refer to the selected study communities as *cities* although, because of the data structure, the counties making up the cities were the actual units of analysis.) Using a hierarchical modeling approach, we have assessed the mortality risks associated with particulate matter (PM) less than 10 µm in aerodynamic diameter (PM_{10}) and the other combustion-related criteria pollutants: ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and carbon monoxide (CO). In the 20-city analysis, we provide detailed findings on the full set of pollutants, combining evidence across the cities to estimate the effects of the individual pollutants while controlling to the extent possible for the effects of the other pollutants. These analyses are then extended to the 90-city database, which we use to explore heterogeneity of effects across broad geographic regions and the determinants of the heterogeneity. The morbidity analysis, which uses hospitalization data from the Health Care Financing Administration (HCFA) for Medicare enrollees, addresses the association of hospitalization risk with PM_{10} and other pollutants in 14 cities. Hierarchical models are used to summarize the effects of air pollution on hospitalization risk. In a separate NMMAPS report, we will provide the findings of a planned joint analysis of morbidity and mortality.

Part I of the NMMAPS Report, *Methods and Methodologic Issues* (Samet et al 2000b), provides comprehensive descriptions of the methods used in the present report to summarize evidence on air pollution and mortality across multiple locations. It also presents a systematic analysis of the problem of measurement error in time-series studies of air pollution and proposes an approach to correcting for the consequences of measurement error, using data from studies with measurements for PM from both personal monitors and central sites. The report also describes 2 conceptually similar analytic approaches to evaluating the extent of associations found in daily time series on short-term mortality displacement (also termed *harvesting*). These methodologic topics, measurement error and mortality displacement, were selected for development because both were proposed as potentially severe limitations to interpretation of the time-series studies.

The objectives for developing specific methodologic components for NMMAPS are fivefold.

1. To develop semiautomated or automated approaches for database construction using databases of the EPA, the National Center for Health Statistics (NCHS), the Health Care Financing Administration (HCFA), the Census Bureau, and the National Weather Service;
2. To develop and apply statistical methods for regression analyses of the multisite data and to develop spatial time-series methods to estimate spatial maps of the relative rates of mortality associated with air pollution, while accounting, as necessary, for the spatial and temporal correlations in the mortality data;
3. To develop and apply methods that adjust for smooth trends and seasonality on mortality caused by changing demographics and health behaviors, influenza epidemics, and other unidentified factors;
4. To examine the consequences of measurement error in the exposure variables for assessing pollutant-mortality associations; and
5. To examine the degree to which pollution-related mortality reduces years of life (mortality displacement).

The objectives for application of methods developed for NMMAPS are threefold.

1. To assess the relation between air pollution and mortality in the largest US cities monitored for PM_{10} from 1987 forward;
2. To assess the relation between air pollution and morbidity in selected US cities monitored for PM_{10} from 1987 forward; and
3. To conduct paired analyses of morbidity and mortality in the same locations.

The design for NMMAPS builds on prior work supported by the Health Effects Institute in the Particle Epidemiology Evaluation Project (PEEP) (Samet et al 1995, 1997). This project was initiated in 1994 with the objectives of validating the data and replicating the findings in several of the time-series studies of air pollution and mortality reported during the 1990s. In a second phase, PEEP addressed several methodologic issues. These included selecting the approach for controlling for potential confounding by weather (Samet et al 1998) and determining the sensitivity of findings to model-building strategies (Kelsall et al 1997; Samet et al 1997).

The present project, NMMAPS, evolved from PEEP. The objectives encompassed methodologic issues that were persistent sources of uncertainty in interpreting the epidemiologic evidence: mortality displacement and exposure measurement error. The plan for multicity analyses was prompted by questioning the rationale for the study locations previously selected and by the prospect of setting this concern aside with analyses conducted using a defined sampling frame. Additionally, advances in hardware and software made this type of analysis feasible. The NMMAPS project was initiated at the end of 1996, as PEEP was ending.

Introduction to NMMAPS Part II

This report provides an integrated synthesis of the key findings of NMMAPS on air pollution and morbidity and mortality. The report begins by introducing the rationale for the multicity approach that is used in NMMAPS and briefly describing the statistical methods used to combine evidence across locations. The findings on mortality are then presented for the 2 databases: the 20 and 90 largest

US cities. In the analysis of the 20 cities, the primary analytic thrust was toward estimating the overall effects of PM_{10} and other criteria air pollutants. We used the previously described Bayesian hierarchical model developed for this purpose (Dominici et al 2000; Samet et al 2000a). Air pollution–mortality associations are assessed within the individual cities with previously described methods (Kelsall et al 1997); the evidence is then combined across the cities using the model of Dominici and coworkers. We next provide the results of using a multistage, regional modeling approach for exploring spatial heterogeneity in the 90-city database. We also evaluate sociodemographic and other characteristics of the cities as determinants of heterogeneity in the effects of PM_{10}.

Hospitalization data are also analyzed by combining information across cities. For morbidity, the cities were selected with preference given to those 14 locations having the most abundant PM_{10} measurements. The within-cities time-series analyses are accomplished with a distributed lag approach developed by Schwartz (2000b). Evidence is then combined across locations using hierarchical methods common to meta-analysis. This approach also allows the examination of sociodemographic characteristics of the population as modifiers of the effect of PM_{10} on heart and lung disease. In addition, the assessment of confounding by other pollutants was done in the second stage of the hierarchical model.

SECTION 1: EXPOSURE MEASUREMENT ERROR IN TIME-SERIES STUDIES OF AIR POLLUTION

(*Scott L Zeger, Duncan Thomas, Francesca Dominici, Jonathan M Samet, Joel Schwartz, Douglas W Dockery, and Aaron Cohen*; reprinted from Samet et al 2000b)

Misclassification of exposure has long been recognized as an inherent limitation of epidemiologic studies of the environment and disease. For many agents of interest, exposures take place over time and in multiple locations; accurately estimating the relevant exposures for an individual participant in epidemiologic studies is often daunting, particularly within the limits set by feasibility, participant burden, and cost. The problem of measurement error is well recognized, and researchers have taken steps to deal with its consequences by limiting the degree of error through the design of a study, estimating the degree of error using a nested validation study, and making adjustments for measurement error in statistical analyses.

Section 1 sets out a systematic conceptual formulation of the problem of measurement error in epidemiologic studies of air pollution and considers the consequences of measurement error within this formulation. When possible, available data were used to make simple estimates of measurement error effects.

The introduction to section 1 presents an overview of the main ideas on measurement errors in linear regression, distinguishing 2 extremes of a continuum: Berkson from classical type errors, and the univariate predictor from the multivariate predictor case. We then propose a single conceptual framework for evaluation of measurement errors in the log-linear regression used for time-series studies of particulate air pollution and mortality, identifying 3 main components of error. We also present new simple analyses of data on exposures of particulate matter less than 10 µm in aerodynamic diameter (PM_{10}) from the Particle Total Exposure Assessment Methodology (PTEAM) study (Ozkaynak et al 1996). Finally, we summarize open questions regarding measurement error and suggest the kind of additional data necessary to address them.

EXPOSURE MEASUREMENT ERROR IN TIME–SERIES STUDIES OF AIR POLLUTION: CONCEPTS AND CONSEQUENCES

(*Scott L Zeger, Duncan Thomas, Francesca Dominici, Jonathan M Samet, Joel Schwartz, Douglas Dockery, and Aaron Cohen*; reprinted from Zeger et al 2000)

Misclassification of exposure is a well-recognized inherent limitation of epidemiologic studies of disease and the environment. For many agents of interest, exposures take place over time and in multiple locations; accurately estimating the relevant exposures for an individual participant in epidemiologic studies is often daunting, particularly within the limits set by feasibility, participant burden, and cost. Researchers have taken steps to deal with the consequences of measurement error by limiting the degree of error through a study's design, estimating the degree of error using a nested validation study, and by adjusting for measurement error in statistical analyses. In this paper, we address measurement error in observational studies of air pollution and health. Because measurement error may have substantial implications for interpreting epidemiologic studies on air pollution, particularly the time–series analyses, we developed a systematic conceptual formulation of the problem of measurement error in epidemiologic studies of air pollution and then considered the consequences within this formulation. When possible, we used available relevant data to make simple estimates of measurement error effects. This paper provides an overview of measurement errors in linear regression, distinguishing two extremes of a continuum-Berkson from classical type errors, and the univariate from the multivariate predictor case. We then propose one conceptual framework for the evaluation of measurement errors in the log-linear regression used for time–series studies of particulate air pollution and mortality and identify three main components of error. We present new simple analyses of data on exposures of particulate matter < 10 µm in aerodynamic diameter from the Particle Total Exposure Assessment Methodology Study. Finally, we summarize open questions regarding measurement error and suggest the kind of additional data necessary to address them.

SECTION 3: MORTALITY DISPLACEMENT–RESISTANT ESTIMATES OF AIR POLLUTION EFFECTS ON MORTALITY (*Scott L Zeger, Francesca Dominici, and Jonathan M Samet*; reprinted from Samet et al 2000c)

A number of studies have recently shown an association between particle concentrations in outdoor air and daily mortality counts in urban locations. In the public health interpretation of this evidence, a key issue is whether the increased mortality associated with higher pollution levels is restricted to very frail persons for whom life expectancy is short in the absence of pollution. This possibility has been termed the *harvesting* or *mortality displacement* hypothesis. We present an approach to estimating the association between pollution and mortality from time-series data that is resistant to short-term mortality displacement. The method is based in the concept that mortality displacement alone creates associations only at shorter time scales. We use frequency domain log-linear regression (FDLLR) to decompose the information about the pollution-mortality association into distinct time scales, and we then create mortality displacement–resistant estimates by excluding the short-term information that is affected by mortality displacement. We illustrate the methods with total suspended particles (TSP) and mortality counts from Philadelphia for 1974 to 1988. We show that the TSP–mortality association in Philadelphia is inconsistent with the mortality displacement–only hypothesis and that the mortality displacement–resistant estimates of the relative risk of mortality associated with TSP are actually larger, not smaller, than the ordinary estimates.

HARVESTING-RESISTANT ESTIMATES OF AIR POLLUTION EFFECTS ON MORTALITY (*Scott L Zeger, Francesca Dominici, and Jonathan Samet*; reprinted from Zeger et al 1999)

A number of studies have recently shown an association between particle concentrations in outdoor air and daily mortality counts in urban locations. In the public health interpretation of this evidence, a key issue is whether the increased mortality associated with higher pollution levels is restricted to very frail persons for whom life expectancy is short in the absence of pollution. This possibility has been termed the "harvesting hypothesis." We present an approach to estimating the association between pollution and mortality from times series data that is resistant to short-term harvesting. The method is based in the concept that harvesting alone creates associations only at shorter time scales. We use frequency domain log-linear regression to decompose the information about the pollution-mortality association into distinct time scales, and we then create harvesting-resistant estimates by excluding the short-term information that is affected by harvesting. We illustrate the methods with total suspended particles and mortality counts from Philadelphia for 1974–1988. The total suspended particles-mortality association in Philadelphia is inconsistent with the harvesting-only hypothesis, and the harvesting-resistant estimates of the total suspended particles relative risk are actually larger—not smaller—than the ordinary estimates.

Affiliations for Authors of Appendix Abstracts (Alphabetical Listing)

Richard T Burnett
Biostatistics and Epidemiology Division
Safe Environments Directorate
Healthy Environments and Consumer Safety Branch
Health Canada

Merlise Clyde
Institute of Statistics and Decisions Sciences
Duke University

Aaron Cohen
Health Effects Institute

Ivan Coursac
PSA–Peugeot–Citroën, France

Frank C Curriero
Department of Biostatistics
School of Hygiene and Public Health
[also Bloomberg School of Public Health]
The Johns Hopkins University

Michael J Daniels
Department of Statistics
Iowa State University

Anup Dewanji
Applied Statistics Unit
Indian Statistical Institute

Douglas Dockery
Department of Environmental Health
Harvard School of Public Health
Harvard University

Francesca Dominici
Department of Biostatistics
School of Hygiene and Public Health
[also Bloomberg School of Public Health]
The Johns Hopkins University

Mark S Goldberg
Department of Medicine and Department of Epidemiology and Biostatistics
McGill University

Trevor J Hastie
Department of Statistics
Stanford University

Daniel Krewski
Department of Epidemiology and Community Medicine
R Samuel McLaughlin Center for Population Risk Assessment
University of Ottawa

Aidan McDermott
Department of Biostatistics
Bloomberg School of Public Health
The Johns Hopkins University

Jonathan M Samet
Department of Epidemiology
School of Hygiene and Public Health
[also Bloomberg School of Public Health]
The Johns Hopkins University

Joel Schwartz
Department of Environmental Health
Harvard School of Public Health
Harvard University

Lianne Sheppard
Departments of Biostatistics and Environmental Health
School of Public Health and Community Medicine
University of Washington

Duncan Thomas
Department of Preventive Medicine
University of Southern California School of Medicine

Antonella Zanobetti
Department of Environmental Health
Harvard School of Public Health
Harvard University

Scott L Zeger
Department of Biostatistics
School of Hygiene and Public Health
[also Bloomberg School of Public Health]
The Johns Hopkins University

ABOUT THE AUTHOR

Francesca Dominici received her PhD in statistics from the University of Padua in 1997. She is an associate professor of biostatistics at The Johns Hopkins Bloomberg School of Public Health. Her research interests include environmental epidemiology, toxicology and carcinogenicity, meta-analysis of clinical trials, estimating health costs of smoking, risk adjustment models, and estimation of group performance in health services and research.

ABBREVIATIONS AND OTHER TERMS

APHENA	Air Pollution and Health—A European and North American Approach
EPA	Environmental Protection Agency (US)
GAM	generalized additive model
LOESS	locally weighted smoothers
NMMAPS	National Morbidity, Mortality, and Air Pollution Study (US)
PM	particulate matter
PM_{10}	PM 10 µm or less in aerodynamic diameter
RJMCMC	Reversible Jump Monte Carlo Markov Chain

INTRODUCTION

In 1998, HEI issued the first Request for Applications (98-5) for the Walter A Rosenblith New Investigator Award (see the Preface to this Research Report). This Award was established to provide funding for up to 3 years for outstanding investigators to conduct independent research at the beginning of their careers. The aim of the Award is to encourage highly qualified and promising investigators to direct their creativity and intellectual curiosity toward evaluating the health effects of air pollution. The Award was named after Professor Rosenblith, the first Chair of the HEI Research Committee, because of his leadership in developing HEI's scientific program and his interest in attracting basic scientists to air pollution research so the field may benefit from their new ideas and fresh approaches.

Evaluating applications includes assessing the applicant's (a) potential to be a leader in the field, (b) institutional research environment and its support, and (c) research proposal for its quality and especially for the originality of its approach.

Francesca Dominici's application clearly met all the qualifications and she became the first recipient of the Rosenblith Award (2000). In this Investigator's Report, Dominici reviews her collaboration with colleagues at The Johns Hopkins University and other academic institutions and chronicles her experience. During the time she was funded by the Rosenblith Award and as a result of the work performed under this contract, she received other honors that also testified to her promise as a young scientist: In 2001, Dominici won the *Young Investigator Award* from the American Statistical Association Section of Statistics in Epidemiology; in 2002, she was the keynote speaker at the joint meeting of the Environmental Statistics Group and General Applications section at the Royal Statistical Society.

SCIENTIFIC BACKGROUND

Since the 1980s, many epidemiologic time-series studies have found associations between short-term increases of fairly low concentrations of ambient particulate matter (PM*) and short-term increases in mortality and morbidity (eg, emergency room visits, hospitalizations; Dockery et al 1993; Schwartz 1993; Pope et al 1995). These studies were generally conducted in single locations chosen for a variety of reasons and were analyzed with different statistical approaches. Many people noted two fundamental questions about these studies: Could pollutants other than PM be responsible for these effects? and Would a study of many locations chosen on the basis of specific criteria and analyzed with uniform methods find similar associations? The National Morbidity, Mortality, and Air Pollution Study (NMMAPS), initiated in 1996 and funded by HEI, was designed to address these and other questions raised about earlier studies.

NMMAPS examined the effect of particulate air pollution on mortality in the 90 largest cities in the United States for which both PM_{10} mass concentrations and specific-cause daily mortality data were available; one analytic approach was applied to all data for all cities (Samet et al 2000b). The levels of PM and other pollutants varied among these locations. Using data from the 90 cities, the investigators explored whether effects of PM_{10} varied across regions of the country. Using morbidity data from 14 of the cities for which daily PM_{10} mass concentrations were available, they evaluated hospitalizations for cardiovascular disease, chronic obstructive lung disease, or pneumonia among individuals 65 years of age and older (Samet et al 2000b). Because of the public health implications of the association between very low concentrations of PM and deleterious health outcomes, NMMAPS evaluated the relation between PM concentrations and mortality in the 20 largest cities (Daniels et al 2004). Most recently, NMMAPS investigators evaluated the association between PM_{10} and morbidity together with mortality in 10 cities that had daily PM_{10} monitoring data available (Dominici et al 2004b).

In the early stages of NMMAPS, the investigators developed methods to address other issues that had been raised about the PM time-series studies: (1) whether pollution data gathered by central monitors are an adequate surrogate for personal exposure (an assumption generally applied in such studies) and the degree to which measurement error affects the results obtained under such assumptions; and (2) whether mortality was advanced by just a few days for frail,

* A list of abbreviations and other terms appears at the end of the Investigator's Report.

near-death individuals (mortality displacement) or by a longer time for other susceptible individuals (Samet et al 2000a). Dr Dominici used funding from the Walter A Rosenblith New Investigator Award to develop methodologic approaches for many of the analyses used in NMMAPS.†

RESEARCH COMPLETED AND RESEARCH PLANNED

RESEARCH OBJECTIVES ADDRESSED BY METHODS DEVELOPED AND IMPLEMENTED

The objectives of the research outlined in Dominici's proposal were based on earlier work conducted with data from cities included in NMMAPS (Samet et al 1995, 1997; Zeger et al 1999). The specific aims proposed were to develop methodologic strategies

1. to assess the short-, medium-, and long-term adverse health effects of air pollution on mortality; and to test the variability of these effects across locations and at different time scales by exploring modeling assumptions;
2. to address the question of whether there is a threshold (beneath which adverse effects cannot be observed) in the shape of a pollution–mortality dose–response curve, and to assess how the shapes of these curves vary across cities; and
3. to establish how the components of measurement error in exposure variables could influence risk assessment.

Dominici proposed to develop and validate more flexible and more general methods to apply to the NMMAPS database to obtain national estimates of air pollution that would be resistant to mortality displacement and to biases from modeling long-term trends inappropriately.

As originally planned, Francesca Dominici, Scott Zeger, Jonathan Samet, and other colleagues addressed these research goals using the NMMAPS data base, as demonstrated by the abstracts of published articles and reports included in Appendix A of the Investigator's Report. They investigated the health effects of pollutants, primarily PM, on cause-specific and all-cause mortality, and conducted extensive sensitivity analyses to determine the time course of health events after high concentrations of air pollutants (lagged effects).

† Dr Dominici's 3-year study, "Air Pollution and Daily Mortality in a National Sampling Frame: Statistical Challenges", began in January 2000. Total expenditures were $286,500. The draft Investigator's Report from Dominici was received for review in January 2004 and was accepted for publication in February 2004. During the review process, the HEI Health Review Committee and the investigator had the opportunity to exchange comments and to clarify issues in both the Investigator's Report and in the Review Committee's Commentary.

To address the first objective described above, Dominici and colleagues extended and implemented methods based on hierarchical Bayesian models to combine city-specific effect estimates and calculated regional or national estimates; these methods accounted for time trends and confounding factors to ensure that the models would capture relevant components (Dominici et al 2000a,b, 2002a, 2003b,d). They applied spatial models to time-series data to estimate and graphically compare the health effects of air pollution across regions (Dominici et al 2003e). They evaluated mortality displacement by applying time-scale Poisson regression models (Zeger et al 1999, 2000a; Dominici et al 2003a). They also improved semiparametric models used to analyze time-series data (Dominici et al 2004a).

To address the second and third objectives, Daniels, Dominici, and colleagues developed Bayesian hierarchical models to evaluate the shape of the relation between air pollution concentrations and mortality (Daniels et al 2000, 2004; Dominici et al 2002a) and how the shapes of this relation might change across cities in the United States.

To address the third objective, Dominici and colleagues developed a multistage Poisson regression model to evaluate how exposure measurement error might affect time-series estimates of PM's effects on mortality. Their approach used (a) a combination of generalized additive models (GAMs) with a log link function and Poisson error for the association between personal exposure and mortality, (b) a multistage linear model to estimate variability across validation datasets in the association between personal exposure and ambient pollutant concentrations, and (c) data augmentation methods to address uncertainty introduced by missing personal-exposure observations in a time-series dataset (Dominici et al 2000c,d, 2003a; Zeger et al 2000b,c).

OTHER CONTRIBUTIONS

In addition to developing the methods originally proposed, Dominici and her colleagues from Harvard University and The Johns Hopkins also developed a two-stage analytic approach to evaluate the effects of air pollution on morbidity together with mortality among elderly residents of the 10 largest cities that had daily monitoring values for PM (Dominici et al 2004b). The two-stage approach incorporated a possible correlation between the effects of PM on mortality rates and on hospitalization rates.

An important aspect of Dominici's work has been the extensive investigation into the choice of models and the effect of various model assumptions on the results of data analysis. This particular investigation led to the discovery that the default settings in the *gam* function in the S-Plus statistical software program in its generally marketed format (Insightful Corp, Seattle WA), which had been used extensively since 1995 by many researchers in the field, were not always appropriate to analyze data in air pollution time-series studies (Dominici et al 2002b). Simultaneously, Canadian researchers found that under certain conditions, the *gam* function underestimated standard errors (Ramsey et al 2003). These findings prompted Dominici and colleagues to develop programming to overcome the limitations in this software program.

Thus, research conducted under this Award has not only advanced methods used to analyze complex time-series data, but has also highlighted the importance of conducting in-depth sensitivity analyses, even when validated statistical methods are being applied.

ANTICIPATED CONTRIBUTIONS FROM ONGOING RESEARCH

Dominici and colleagues are continuing to explore, develop, and improve methods to analyze time-series data in studies of air pollution and health. Ongoing projects include (1) evaluating improved semiparametric models for time-series analyses of air pollution and mortality; (2) evaluating the interactions among seasons, air pollutant concentrations, and other confounders; (3) evaluating bias in exposure–response analyses based on spatial–temporal aggregation; and (4) developing Bayesian hierarchical and distributed lag models to determine the association between ozone and mortality from cardiovascular conditions in 19 cities (data from 1987–1994) and between ozone and total (all-cause) mortality in 95 cities (data from 1987–2000). They are also developing approaches to estimate distributed lag models for time-series analyses of air pollution and mortality to use when some air pollution data are missing.

Data from the National Medicare Cohort Air Pollution Study will allow Samet, Dominici, Zeger, and their colleagues to continue to explore the effects of PM and other pollutants on morbidity and mortality simultaneously, as well as the joint effects of fine and ultrafine PM and other pollutants on persons both with and without underlying cardiac or pulmonary disease.

The Air Pollution and Health—A European and North American Approach (APHENA), a study funded jointly by HEI and the European Commission, is evaluating not only the effects of pollutants, but especially the heterogeneity of effects among regions with different sources of pollutants and different frequencies of pollutant monitoring. Dominici and colleagues are collaborating closely with other APHENA investigators to explore, compare, and further improve analytic methods to evaluate the effects of air pollutants on morbidity and mortality across cities, regions, countries, and continents.

With HEI funding, Dominici and colleagues are developing the internet-based Health and Air Pollution Surveillance System, which will allow other researchers to access and analyze US urban data related to air pollution, morbidity, and mortality.

DISCUSSION

With large data sets becoming widely available, the potential for addressing many questions related to the health effects of air pollution is enhanced. The unique methodologic challenges raised require reliable statistical and epidemiologic strategies to be developed and implemented. Methods of analysis from time-series studies, biostatistics, quantitative epidemiology, and machine learning are being combined in new ways to evaluate these data; hence, a combination of theoretical analysis and simulation is essential for assessing the properties of the newly developed methods. HEI funding has fostered such explorations with the Rosenblith Award and with other projects (eg, Navidi et al 1999).

As a recipient of the Award, Dominici was able to devote more time to developing methods and collaborating with other researchers at The Johns Hopkins University and other US and international institutions, some of whom also had received funding from HEI for their projects.

An unexpected result of her work was the discovery and announcement by her team of the shortcomings of the *gam* function in S-Plus statistical software (Insightful Corp, Seattle WA). This finding came at a point when NMMAPS and other time-series studies were being assessed by the US Environmental Protection Agency (EPA) as part of its periodic review of the National Ambient Air Quality Standards for PM; thus, this discovery led to considerable public and academic interest. Dominici quickly assumed leadership for publicizing the nature of the problem and its implications; designing a web page to share progress in resolving the issue; and working with Trevor Hastie at Stanford University, the author of the software, to correct the programming so it would provide more accurate analyses of air pollution time-series data (Dominici et al 2002b, 2004a).

To meet the EPA's deadline, this finding required extensive new analyses of NMMAPS data, which had been previously analyzed with the *gam* default programming, to be carried out and published immediately.

In spite of these complications, Dominici successfully completed all the research outlined in her proposal. She published several articles in excellent peer-reviewed journals, and the NMMAPS team has contributed to several HEI Research Reports (Samet et al 2000a,b; Daniels et al

2004; Dominici et al 2004b) and a Special Report (Dominici et al 2003a,b).

As Dominici recounts in the Preface to this Investigator's Report, when she received the Rosenblith Award she was pleased by the significant message it communicated: that developing analytic methods is recognized as making an important contribution in the field of air pollution and health research. Indeed, the work she has accomplished as recipient of this Award reinforces the importance of continuing to develop, evaluate, and improve analytic methods in this field; as the complexity of available data allows more in-depth evaluation of factors known or suspected to be related to or to influence the effects of air pollution on the health of populations, the complexity and accuracy of analytic methods must keep pace.

As the Investigator's Report makes clear, Dominici will continue to contribute to the field with many projects under way that are related to the research funded under this Award, both in methods development and in data analysis. For example, a much more expanded data collection effort is under way, with plans to make the data readily available to other researchers via the web; and she continues with a research program to develop methods for analyzing complex datasets.

Fulfilling the goal of the Rosenblith Award, Francesca Dominici has developed from a promising researcher into a leader in the field of statistical methods for research on the health effects of air pollution. She and her colleagues have advanced our understanding in this area by developing and applying complex statistical methods to describe the relation between air pollution concentrations and morbidity and mortality from specific causes.

ACKNOWLEDGMENTS

The Health Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigator's Report. 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 Virgi Hepner for providing science editing; and to Melissa R Harke, Frederic R Howe, Jenny Lamont, Kasey L Oliver, and Ruth E Shaw for their roles in publishing this Research Report.

REFERENCES

Daniels MJ, 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 152:397–406.

Daniels MJ, Dominici F, Zeger SL, Samet JM. 2004. The National Morbidity, Mortality, and Air Pollution Study, Part III: PM_{10} Concentration–Response Curves and Thresholds for the Largest 20 US Cities. Research Report 94. Health Effects Institute, Boston MA.

Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG Jr, Speizer FE. 1993. An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759.

Dominici F, Daniels M, McDermott A, Zeger SL, Samet JM. 2003a. Shape of the exposure-response relation and mortality displacement in the NMMAPS database. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report, pp 91–96. Health Effects Institute, Boston MA.

Dominici F, Daniels M, Zeger SL, Samet JM. 2002a. 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. 2003b. 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, McDermott A, Hastie T. 2004a. Improved semiparametric time series models of air pollution and mortality. J Am Stat Assoc (in press).

Dominici F, McDermott A, Zeger SL, Samet JM. 2002b. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol 156:193–203.

Dominici F, McDermott A, Zeger SL, Samet JM. 2003d. Airborne particulate matter and mortality: Timescale effects in four US cities. Am J Epidemiol 157:1055–1065.

Dominici F, McDermott A, Zeger SL, Samet JM. 2003e. National maps of the effects of particulate matter on mortality: Exploring geographical variation. Environ Health Perspect 111:39–43.

Dominici F, Samet JM, Zeger SL. 2000a. Combining evidence on air pollution and daily mortality from twenty largest US cities: A hierarchical modeling strategy (with discussion). Royal Stat Soc Ser A 163:263–276.

Dominici F, Samet JM, Zeger SL. 2000b. Combining evidence on air pollution and daily mortality from twenty largest US cities. In: The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues. Research Report 94, pp 61–74. Health Effects Institute, Cambridge MA.

Dominici F, Zanobetti A, Zeger SL, Schwartz J, Samet JM. 2004b. The National Morbidity, Mortality, and Air Pollution Study, Part IV: Hierarchical Bivariate Time-Series Models. Research Report 94. Health Effects Institute, Boston MA. In press.

Dominici F, Zeger SL, Samet JM. 2000c. A measurement error model for time-series studies of air pollution and mortality. Biostatistics 1:157–175.

Dominici F, Zeger SL, Samet JM. 2000d. A measurement error model for time-series studies of air pollution and mortality. In: The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues. Research Report 94, pp 29–41. Health Effects Institute, Cambridge MA.

Health Effects Institute. 2003. Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. Health Effects Institute, Boston MA.

Navidi W, Thomas D, Langholz B, Stram D. 1999. Statistical Methods for Epidemiologic Studies of the Health Effects of Air Pollution. Research Report 86. Health Effects Institute, Cambridge MA.

Pope CA III, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW Jr. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir Crit Care Med 151:669–674.

Ramsay TO, Burnett RT, Krewski D. 2003. The effect of concurvity in generalized additive models linking mortality to ambient particulate matter. Epidemiology 14:18–23.

Samet JM, Dominici F, McDermott A, Zeger SL. 2003. New problems for an old design: Time series analyses of air pollution and health. Epidemiology 14:11–12.

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, 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. 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.

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. Health Effects Institute, Cambridge.

Schwartz J. 1993. Particulate air pollution and chronic respiratory disease. Environ Res 62:7–13.

Zeger SL, Dominici F, Samet JM. 1999. Harvesting-resistant estimates of air pollution effects on mortality. Epidemiology 10:171–175.

Zeger SL, Dominici F, Samet JM. 2000a. Mortality displacement–resistant estimates of air pollution effects on mortality. In: The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues. Research Report 94, pp 43–49. Health Effects Institute, Cambridge MA.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, Cohen A. 2000b. Exposure measurement error in time-series studies of air pollution: Concepts and consequences. Environ Health Perspect 108:419–426.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery DW, Cohen AJ. 2000c. Exposure measurement error in time-series studies of air pollution. In: The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodologic Issues. Research Report 94, pp 15–27. Health Effects Institute, Cambridge MA.

RELATED HEI PUBLICATIONS: PARTICULATE MATTER

	Title	Principal Investigator	Date*
Research Reports			
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		
	Part I. Methods and Methodologic Issues	JM Samet	2000
	Part II. Morbidity and Mortality from Air Pollution in the United States	JM Samet	2000
	Part III. PM_{10} Concentration–Response Curves and Thresholds for the 20 Largest US Cities	MJ Daniels	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 Effect of Air Pollution	W Navidi	1999
75	Ozone Exposure and Daily Mortality in Mexico City: A Time-Series Analysis	DP Loomis	1996
Special Reports			
	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		
	Phase I.A. Replication and Validation of Selected Studies		1995
	Phase I.B. Analyses of the Effects of Weather and Multiple Air Pollutants		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.

HEALTH
EFFECTS
INSTITUTE

Charlestown Navy Yard
120 Second Avenue
Boston MA 02129-4533 USA
+1-617-886-9330
www.healtheffects.org

RESEARCH
REPORT

Number 123
December 2004