RESEARCH REPORT

HEALTH EFFECTS INSTITUTE

Number 168
November 2012

Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments

Richard D. Morgenstern, Winston Harrington, Jhih-Shyang Shih, and Michelle L. Bell

Includes a Commentary by the Institute's Health Review Committee

Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments

Richard D. Morgenstern, Winston Harrington,
Jhih-Shyang Shih, and Michelle L. Bell

with a Commentary by the HEI Health Review Committee

Research Report 168
Health Effects Institute
Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

Publishing history: This document was posted at *www.healtheffects.org* in November 2012.

Citation for document:

Morgenstern RD, Harrington W, Shih J-S, Bell ML. 2012. Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments. Research Report 168. Health Effects Institute, Boston, MA.

 Asterisk Typographics, Barre, Vt., Compositor. Printed by Recycled Paper Printing, Boston, Mass. Library of Congress Catalog Number for the HEI Report Series: WA 754 R432.

Cover paper: made with at least 55% recycled content, of which at least 30% is post-consumer waste; free of acid and elemental chlorine. Text paper: made with 100% post-consumer waste recycled content; acid free; no chlorine used in processing. The book is printed with soy-based inks and is of permanent archival quality.

CONTENTS

Research Report 168

ABOUT HEI

The Health Effects Institute is a nonprofit corporation chartered in 1980 as an independent research organization to provide high-quality, impartial, and relevant science on the effects of air pollution on health. To accomplish its mission, the institute

- Identifies the highest-priority areas for health effects research;
- Competitively funds and oversees research projects;
- Provides intensive independent review of HEI-supported studies and related research;
- Integrates HEI's research results with those of other institutions into broader evaluations; and
- Communicates the results of HEI's research and analyses to public and private decision makers.

HEI typically receives half of its core funds from the U.S. Environmental Protection Agency and half from the worldwide motor vehicle industry. Frequently, other public and private organizations in the United States and around the world also support major projects or research programs. HEI has funded more than 280 research projects in North America, Europe, Asia, and Latin America, the results of which have informed decisions regarding carbon monoxide, air toxics, nitrogen oxides, diesel exhaust, ozone, particulate matter, and other pollutants. These results have appeared in the peer-reviewed literature and in more than 200 comprehensive reports published by HEI.

HEI's independent Board of Directors consists of leaders in science and policy who are committed to fostering the public–private partnership that is central to the organization. The Health Research Committee solicits input from HEI sponsors and other stakeholders and works with scientific staff to develop a Five-Year Strategic Plan, select research projects for funding, and oversee their conduct. The Health Review Committee, which has no role in selecting or overseeing studies, works with staff to evaluate and interpret the results of funded studies and related research.

All project results and accompanying comments by the Health Review Committee are widely disseminated through HEI's Web site (*www.healtheffects.org*), printed reports, newsletters and other publications, annual conferences, and presentations to legislative bodies and public agencies.

ABOUT THIS REPORT

Research Report 168, *Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments*, presents a research project funded by the Health Effects Institute and conducted by Dr. Richard D. Morgenstern of Resources for the Future, Washington, D.C., and his colleagues. This report contains three main sections.

The HEI Statement, prepared by staff at HEI, is a brief, nontechnical summary of the study and its findings; it also briefly describes the Health Review Committee's comments on the study.

The Investigators' Report, prepared by Morgenstern and colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

The Commentary is prepared by members of the Health Review Committee with the assistance of HEI staff; it places the study in a broader scientific context, points out its strengths and limitations, and discusses remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI's rigorous review process. When an HEI-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. This draft report is first examined by outside technical reviewers and a biostatistician. The report and the reviewers' comments are then evaluated by members of the Health Review Committee, an independent panel of distinguished scientists who have no involvement in selecting or overseeing HEI studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Commentary reflects the information provided in the final version of the report.

PREFACE

HEI's Outcomes Research Program

The goal of most air quality regulations is to protect the public's health by implementing regulatory actions or providing economic incentives that help reduce the public's exposure to air pollutants. If this goal is met, air pollution should be reduced, and indicators of public health should improve or at least not deteriorate. Evaluating the extent to which air quality regulations succeed in protecting public health is part of a broader effort — variously termed *outcomes research*, *accountability research*, or *research on regulatory effectiveness* — designed to assess the performance of environmental regulatory policies in general. In recent decades, air quality in the United States and Western Europe has improved substantially, and this improvement is attributable to a number of factors, including increasingly stringent air quality regulations. However, the cost of the pollution-control technologies and mechanisms needed to implement and enforce these regulations is often high. It is therefore prudent to ask whether the regulations have in fact yielded demonstrable improvements in public health, which will provide useful feedback to inform future efforts.

Several U.S. government agencies have concluded that direct evidence about the extent to which air quality regulations have improved health (measured as a decrease in premature mortality and excess morbidity) is lacking. This finding is well documented by the National Research Council (NRC) in its report *Estimating the Public Health Benefits of Proposed Air Pollution Regulations* (NRC 2002), as well as by the California Air Resources Board, the U.S. Environmental Protection Agency (EPA), the U.S. Centers for Disease Control and Prevention (CDC), and other agencies.

In 2003, the Health Effects Institute published a monograph on outcomes research, Communication 11, *Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research* (HEI 2003). This monograph was written by the members of HEI's multidisciplinary Accountability Working Group after a 2001 workshop on the topic. Communication 11 set out a conceptual framework for outcomes research and identified the types of evidence required and the methods by which the evidence should be obtained. It has also guided the development of the HEI Health Outcomes Research program, which is discussed below.

Between 2002 and 2004, HEI issued four requests for applications (RFAs) for studies to evaluate the effects of actions taken to improve air quality. The study by Dr. Richard Morgenstern and colleagues described in this Research Report (Morgenstern et al. 2012) was funded under RFA 04-4, "Measuring the Health Impact of Actions Taken to Improve Air Quality." HEI funded eight additional outcomes studies resulting from this and other RFAs (see Preface Table).

This preface describes both the framework of outcomes research as it relates to air quality regulations and HEI's Outcomes Research program.

BACKGROUND

The first step in assessing the effectiveness of air quality regulations is to measure emissions of the targeted pollutants to see whether they have in fact decreased as intended. A series of intermediate assessments, described in detail below, are needed in order to accurately measure the adverse health effects associated with air pollution to see whether they, too, decreased in incidence or severity relative to emissions. Some outcomes studies to date have used hypothetical scenarios (comparing estimated outcomes under existing and more stringent regulations) and risk estimates obtained from epidemiologic studies in an attempt to quantify past effects on health and to predict future effects (U.S. EPA 1999). However, more extensive validation of these estimates with data on actual outcomes would be helpful.

The long-term improvements in U.S. air quality have been associated with improved health in retrospective epidemiologic studies (Chay and Greenstone 2003;

HEI's Outcomes Research Program[a]

RFA / Investigator (Institution)	Study or Report Title	Intervention
RFA 02-1		
Douglas Dockery (Harvard School of Public Health, Boston, MA)	"Effects of Air Pollution Control on Mortality and Hospital Admissions in Ireland" (in press)	Coal ban in Irish cities
Annette Peters (GSF–National Research Center for Environment and Health, Neuherberg, Germany[b])	The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany (published as Research Report 137, 2009)	Switch from brown coal to natural gas for home heating and power plants, changes in motor vehicle fleet after reunification of Germany
RFA 04-1		
Frank Kelly (King's College London, London, U.K.)	The Impact of the Congestion Charging Scheme on Air Quality in London: Part 1. Emissions Modeling and Analysis of Air Pollution Measurements. Part 2. Analysis of the Oxidative Potential of Particulate Matter (published as Research Report 155, 2011)	Measures to reduce traffic congestion in the inner city of London
RFA 04-4		
Frank Kelly (King's College London, London, U.K.)	The London Low Emission Zone Baseline Study (published as Research Report 163, 2011)	Measures to exclude most polluting vehicles from entering greater London
Richard Morgenstern (Resources for the Future, Washington, DC)	Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments (published as Research Report 168, 2012)	Measures to reduce sulfur emissions from power plants east of the Mississippi River
Curtis Noonan (University of Montana, Missoula, MT)	Assessing the Impact of a Wood Stove Replacement Program on Air Quality and Children's Health (published as Research Report 162, 2011)	Woodstove change-out program
Jennifer Peel (Colorado State University, Fort Collins, CO)	Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes (published as Research Report 148, 2010)	Measures to reduce traffic congestion during the Atlanta Olympics
Chit-Ming Wong (University of Hong Kong, Hong Kong)	Impact of the 1990 Hong Kong Legislation for Restriction on Sulfur Content in Fuel (published as Research Report 170, 2012)	Measures to reduce sulfur content in fuel for motor vehicles and power plants
RFPA 05-3		
Junfeng (Jim) Zhang (University of Medicine and Dentistry of New Jersey, Piscataway, NJ)	"Molecular and Physiological Responses to Drastic Changes in PM Concentration and Composition" (in press)	Measures to improve air quality during the Beijing Olympics

[a] Abbreviations: RFA, Request for Applications; RFPA, Request for Preliminary Applications.

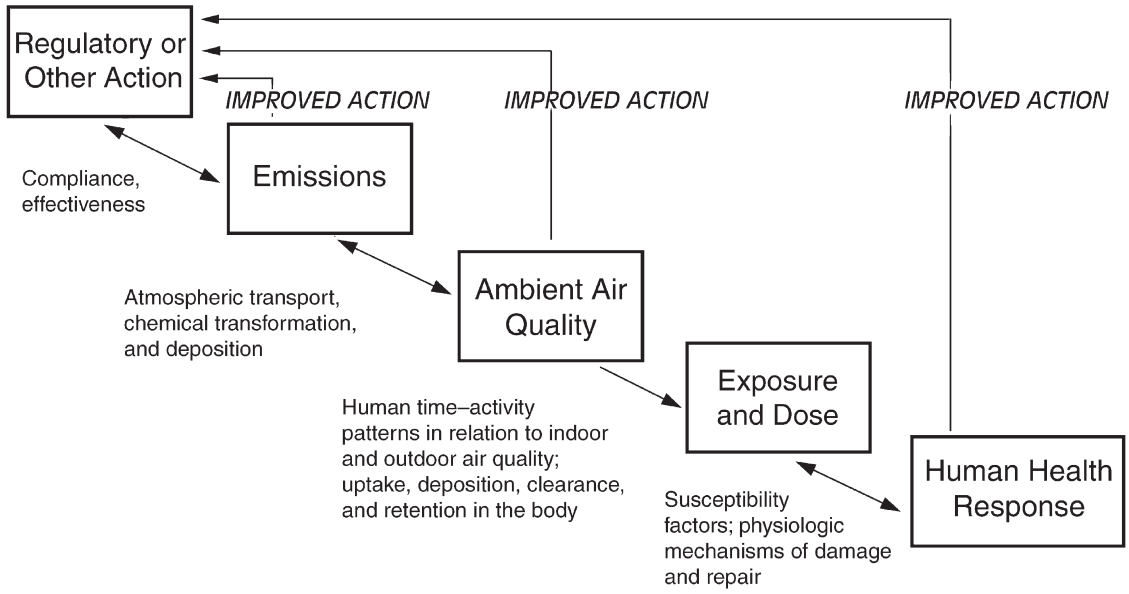
[b] As of 2008, this institution has been called the Helmholtz Zentrum München–German Research Center for Environmental Health.

Laden et al. 2006; Pope et al. 2009). Considerable challenges, however, are inherent in the assessment of the health effects of air quality regulations. Different regulations go into effect at different times, for example, and may be implemented at different levels of government (e.g., national, regional, or local). Their effectiveness therefore needs to be assessed in ways that take into account the varying times of implementation and levels of regulation. In addition, other changes at the same time and place might confound an apparent association between pollution reduction and improved health, such as economic trends (e.g., changes in employment), improvements in health care, and behavioral changes (e.g., staying indoors when government warnings indicate pollution concentrations are high). Moreover, adverse health effects that might have been caused by exposure to air pollution can also be caused by other environmental risk factors (some of which may have changed over the same time periods as the air pollution concentrations). These challenges become more pronounced when regulations are implemented over long periods and when changes in air quality and health outcomes are not seen immediately, thus increasing the chance for confounding by other factors. For these reasons, scenarios in which regulations are expected to have resulted in rapid changes in air quality tend to be among the first, and most likely, targets for investigation, rather than evaluations of complex regulatory programs implemented over multiple years. Studies in Ireland by Clancy and colleagues (2002) and in Hong Kong by Hedley and colleagues (2002) are examples of such scenarios.

These inherent challenges are well documented in Communication 11 (HEI 2003), which was intended to advance the concept of outcomes research and to foster the development of methods and studies throughout the relevant scientific and policy communities. In addition, recent advances in data collection and analytic techniques provide an unprecedented opportunity to improve our assessments of the effects of air quality interventions.

THE OUTCOMES EVALUATION CYCLE

The NRC's Committee on Research Priorities for Airborne Particulate Matter set out a conceptual framework for linking air pollution sources to adverse health effects (NRC 1998). This framework can be used to identify factors along an "outcomes evaluation cycle" (see Figure below), each stage of which affords



Outcomes Evaluation Cycle. Each box represents a stage in the process between regulatory action and human health responses to air pollution. Arrows connecting the stages indicate possible directions of influence. The text below the arrows identifies factors affecting the effectiveness of regulatory actions at each stage. At several of the stages, knowledge gained from studies on outcomes can provide valuable feedback for improving regulatory or other actions.

its own opportunities for making quantitative measurements of the intended improvements.

At the first stage (regulatory action), one can assess whether controls on source emissions have in fact been put into place. At the second stage (emissions), one can determine whether controls on sources have indeed reduced emissions, whether emitters have changed their practices, and whether there have been unintended consequences. At the third stage (ambient air quality), one can assess whether controls on sources and reductions in emissions have resulted in improved air quality. At the fourth stage (personal or population exposure), one can assess whether the improvement in air quality has reduced people's actual exposure and whether susceptible subpopulations (those most likely to experience adverse health effects) have benefited. At this stage, it is important to take into account changes in time–activity patterns that could either increase or reduce exposure. The actual dose that an individual's organs may be exposed to should also be considered (i.e., whether reductions in exposure have led to reductions in concentrations in body tissues such as the lung). Finally, at the fifth stage (human health response), one can assess whether risks to health have declined, given the evidence about changes in health outcomes such as morbidity and mortality that have resulted from changes in exposure. The challenge at this stage is to investigate the health outcomes that are most directly related to exposure to air pollution.

At each stage in the outcomes evaluation cycle, the opportunity exists to collect evidence that either validates the assumptions that motivated the intervention or points to ways in which the assumptions were incorrect. The collection of such evidence can thus ensure that future interventions are maximally effective.

Ultimately, the framework for outcomes research will need to encompass investigations of the broader consequences of regulations, not just the intended consequences. Unintended consequences should also be investigated, along with the possibility that risks to public health in fact increased, as discussed by Wiener (1998) and others who have advanced the concept of a portfolio of effects of a regulation.

HEI'S OUTCOMES RESEARCH PROGRAM

HEI's Outcomes Research program currently includes nine studies. The study by Dr. Richard Morgenstern and colleagues presented in this report is the seventh to be published. The remaining studies are in press and are expected to be published in 2013.

These studies involve the measurement of indicators along the entire outcomes evaluation cycle, from regulatory or other interventions to human health outcomes. Some of the studies focused on interventions that are implemented over relatively short periods of time, such as a ban on the sale of coal, the replacement of old wood stoves with more efficient, cleaner ones, reductions in the sulfur content of fuels, and measures to reduce traffic. Other groups focused on longer-term, wider-ranging interventions or events; for instance, one study assessed complex changes associated with the reunification of the former East and West Germany, including a switch from brown coal to natural gas for fueling power plants and home-heating systems and an increase in the number of modern diesel-powered vehicles in eastern Germany. HEI is also supporting research, including the development of methods, in an especially challenging area, namely, assessment of the effects of regulations implemented incrementally over extended periods of time, such as those, examined in the current study, that resulted from Title IV of the 1990 Clean Air Act Amendments (U.S. EPA 1990), which aimed at reducing sulfur dioxide emissions from power plants by requiring compliance with prescribed emission limitations. Studies on health outcomes funded by HEI to date are summarized in the Preface Table on page x and described in more detail in an interim evaluation of the HEI Outcomes Research program (van Erp and Cohen 2009; van Erp et al. 2012).

FUTURE DIRECTIONS

As a part of its Strategic Plan for 2010 through 2015 (HEI 2010a), HEI has looked closely at opportunities for unique new contributions to health outcomes research. Key recommendations for future research were made at a December 2009 planning workshop (HEI 2010b), which led to HEI issuing a new Request for Applications in January 2011 for a second wave of outcomes research. RFA 11-1, "Health Outcomes Research — Assessing the Health Outcomes of Air Quality Actions," solicited applications for studies designed to assess the health effects of actions to improve air quality and to develop methods required for, and specifically suited to, conducting such research.

Recently, HEI approved four studies: two will evaluate regulatory and other actions at the national or regional level implemented over multiple years; a third study will evaluate complex sets of actions targeted at improving air quality in large urban areas and major ports with well-documented air quality problems and programs to address them; and a fourth study will develop methods to support such health outcomes research. These studies are currently underway.

In addition, HEI has funded the development of two Web sites intended to enhance transparency and provide other researchers with access to extensive data and software from HEI-funded studies:

1. Data and software from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), as described by Zeger and colleagues (2006) (data available at the Johns Hopkins Bloomberg School of Public Health Web site *www.ihapss.jhsph.edu*); and
2. Data from the National Particle Component Toxicity Initiative (NPACT) on concentrations of components of particulate matter with an aerodynamic diameter $\leq$ 2.5 μm ($PM_{2.5}$) collected at or near the 54 sites in the EPA's $PM_{2.5}$ Chemical Speciation Trends Network (STN) (data available at the Atmospheric and Environmental Research Web site *https://hei.aer.com*).

The data on pollution and health from a large number of U.S. cities, as documented by the NMMAPS team and made available on the Internet-Based Health and Air Pollution Surveillance System (iHAPSS) Web site, constitute a valuable resource that allows other researchers to undertake additional analyses, possibly including further outcomes studies. The STN Web site provides scientists an opportunity to investigate specific questions about concentrations of $PM_{2.5}$ components and their association with adverse health effects in regions covered by the STN network and to address questions related to outcomes research when interventions in these regions are being planned.

In January 2008, HEI co-organized and cosponsored, with the CDC's National Environmental Public Health Tracking Program and the EPA, a workshop titled "Methodologic Issues in Environmental Public Health Tracking of Air Pollution Effects." The workshop was part of an effort to implement the initiative outlined in HEI's Strategic Plan for 2005 through 2010 (HEI 2005) to "build networks with the U.S. Centers for Disease Control and Prevention and state public health tracking programs to facilitate accountability research."

The workshop built on the work of the CDC's National Environmental Public Health Tracking Program (see the CDC Web site *www.cdc.gov/nceh/tracking/*) in the development of standardized measures of air pollution–related effects on health at the state and local levels in the United States. It brought together representatives of state and federal agencies and academic researchers to discuss methodologic issues in developing standardized measures and made recommendations for their further development and application in assessing the health impacts of air pollution, including the impacts of actions taken to improve air quality. The recommendations were provided in a September 2008 report to the CDC, and the proceedings were published in the journal *Air Quality, Atmosphere & Health* in December 2009 (Matte et al. 2009). The CDC has subsequently funded a pilot project under the National Environmental Public Health Tracking Program to implement the recommendations of the workshop in selected states and metropolitan areas.

HEI will continue to seek opportunities to work with the CDC and the EPA to apply methods newly developed for tracking public health and assessing the effectiveness of environmental regulations.

Investigators who have identified a distinctive opportunity to evaluate the effects of environmental regulations on air pollution and human health are encouraged to contact HEI.

REFERENCES

Atmospheric and Environmental Research, Inc. (San Ramon, CA). HEI Air Quality Database. *https://hei.aer.com.* Accessed 10/19/11.

Chay KY, Greenstone M. 2003. The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. Q J Econ 118:1121–1167.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air-pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Health Effects Institute. 2003. Assessing Health Impact of Air Quality Regulations: Concepts and Methods for

Accountability Research. Communication 11. Health Effects Institute, Boston, MA.

Health Effects Institute. 2005. HEI Strategic Plan for Understanding Health Effects of Air Pollution 2005–2010. Health Effects Institute, Boston, MA.

Health Effects Institute. 2010a. HEI Strategic Plan for Understanding the Health Effects of Air Pollution 2010–2015. Health Effects Institute, Boston, MA.

Health Effects Institute. 2010b. Proceedings of an HEI Workshop on Further Research to Assess the Health Impacts of Actions Taken to Improve Air Quality. Communication 15. Health Effects Institute, Boston, MA.

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. 2002. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: An intervention study. Lancet 360:1646–1652.

Johns Hopkins Bloomberg School of Public Health (Baltimore, MD). Internet-Based Health and Air Pollution Surveillance System (last updated 3/19/05). *www.ihapss.jhsph.edu.* Accessed 10/19/11.

Kelly F, Anderson HR, Armstrong B, Atkinson R, Barratt B, Beevers S, Derwent D, Green D, Mudway I, Wilkinson P. 2011a. Part 1. Emissions modeling and analysis of air pollution measurements. In: The Impact of the Congestion Charging Scheme on Air Quality in London. Research Report 155. Health Effects Institute, Boston, MA.

Kelly F, Anderson HR, Armstrong B, Atkinson R, Barratt B, Beevers S, Derwent D, Green D, Mudway I, Wilkinson P. 2011b. Part 2. Analysis of the oxidative potential of particulate matter. In: The Impact of the Congestion Charging Scheme on Air Quality in London. Research Report 155. Health Effects Institute, Boston, MA.

Kelly FJ, Armstrong B, Atkinson R, Anderson HR, Barratt B, Beevers S, Cook D, Green D, Derwent D, Mudway I, Wilkinson P. 2011c. The London Low Emission Zone Baseline Study. Research Report 163. Health Effects Institute, Boston, MA.

Laden F, Schwartz J, Speizer FE, Dockery DW. 2006. Reduction in the particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study. Am J Respir Crit Care Med 173:667–672.

Matte TD, Cohen A, Dimmick F, Samet J, Sarnat J, Yip F, Jones N. 2009. Summary of the workshop on methodologies for environmental public health tracking of air pollution effects. Air Qual Atmos Health 2: 177–184.

Morgenstern RD, Harrington W, Shih J-S, Bell ML. 2012. Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments. Research Report 168. Health Effects Institute, Boston, MA.

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

National Research Council (U.S.). 2002. Estimating the Public Health Benefits of Proposed Air Pollution Regulations. National Academy Press, Washington, DC.

Noonan CW, Ward TJ, Navidi W, Sheppard L, Bergauff M, Palmer C. 2011. Assessing the Impact of a Wood Stove Replacement Program on Air Quality and Children's Health. Research Report 162. Health Effects Institute, Boston, MA.

Peel JL, Klein M, Flanders WD, Mulholland JA, Tolbert PE. 2010. Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes. Research Report 148. Health Effects Institute, Boston, MA.

Peters A, Breitner S, Cyrys J, Stölzel M, Pitz M, Wölke G, Heinrich J, Kreyling W, Küchenhoff H, Wichmann H-E. 2009. The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany. Research Report 137. Health Effects Institute, Boston, MA.

Pope CA III, Ezzati M, Dockery DW. 2009. Fine-particulate air pollution and life expectancy in the United States. N Engl J Med 360:376–386.

U.S. Centers for Disease Control and Prevention. National Environmental Public Health Tracking Program. *www.cdc.gov/nceh/tracking/*. Accessed 10/19/11.

U.S. Environmental Protection Agency. 1990. A Bill to Amend the Clean Air Act to Provide for Attainment and Maintenance of Health Protective National Ambient Air Quality Standards, and for Other Purposes. S 1630, 101st Cong, 2nd Sess.

U.S. Environmental Protection Agency. 1999. Benefits and Costs of the Clean Air Act 1990 to 2010: Report

to Congress. EPA/410/R-99-001. Office of Air and Radiation, Washington, DC.

van Erp AM, Cohen AJ. 2009. HEI's Research Program on the Impact of Actions to Improve Air Quality: Interim Evaluation and Future Directions. Communication 14. Health Effects Institute, Boston, MA.

van Erp AM, Cohen AJ, Shaikh R, O'Keefe R. 2012. Recent progress and challenges in assessing the effectiveness of air quality interventions toward improving public health: The HEI experience. EM October: 22–28.

Wiener J. 1998. Managing the iatrogenic risks of risk management. Risk Health Safety Environ 9:39–82.

Wong C-M, Rabl A, Thach TQ, Chau YK, Chan KP, Cowling BJ, Lai HK, Lam TH, McGhee SM, Anderson HR, Hedley AJ. 2012. Impact of the 1990 Hong Kong Legislation for Restriction on Sulfur Content in Fuel. Research Report 170. Health Effects Institute, Boston, MA.

Zeger SL, McDermott A, Dominici F, Peng R, Samet J. 2006. Internet-Based Health and Air Pollution Surveillance System. Communication 12. Health Effects Institute, Boston, MA.

HEI STATEMENT

Synopsis of Research Report 168

Evaluating the Effects of Title IV of the 1990 Clean Air Act Amendments on Air Quality

BACKGROUND

As our understanding of the adverse health effects associated with exposure to particulate matter (PM) ≤ 2.5 µm in aerodynamic diameter ($PM_{2.5}$) has grown, ambient $PM_{2.5}$ concentrations have been increasingly regulated in the United States and Europe. Although it is desirable to verify that air quality regulations have resulted in improved air quality, improved health, and reduced mortality, evidence for verification — particularly in terms of health outcomes — has not often been systematically collected and is difficult to establish retrospectively.

In 2000, HEI launched an initiative to improve the evidentiary and methodologic bases for assessing the health impact of regulations and other actions or situations resulting in improved air quality; since then it has funded nine studies through four Requests for Applications (RFAs) issued between 2002 and 2004. The current study, led by Dr. Richard D. Morgenstern of Resources for the Future, was funded under RFA 04-4, which sought proposals for studies of the health effects associated with planned actions to improve air quality (or other situations resulting in marked air quality improvements). Morgenstern and his team analyzed the effects of reductions in pollutants from power plants on $PM_{2.5}$ concentrations in the eastern United States between 1999 and 2005 using a statistical model linking emissions and air quality monitoring data.

Title IV of the Clean Air Act of 1990, entitled "Acid Deposition Control," called for a permanent 10-million-ton reduction in sulfur dioxide (SO_2) emissions from 1980 levels and required installation of continuous monitoring equipment for SO_2 emissions to ensure compliance and track improvements. Beginning with the 1997 National Ambient Air Quality Standards for PM, the U.S. Environmental Protection Agency (EPA) set standards for $PM_{2.5}$ concentrations, and $128 million was appropriated for a nationwide array of $PM_{2.5}$ monitoring stations known as the Air Quality System monitoring network.

When evaluating a regulatory action intended to improve air quality, either prospectively or retrospectively, EPA scientists frequently employ a chemical transport model, such as the Community Multiscale Air Quality (CMAQ) modeling system. One key limitation to the use of such models is that they are based on modeling estimations and not monitoring data. For the current study, Morgenstern and his team proposed a novel model that was data-driven, in that it depended on measured values of emissions and pollutant concentrations and was inherently observational. The statistical models that the investigators developed to link changes in emissions of SO_2 and nitrogen oxides (NO_x) to changes in ambient $PM_{2.5}$ concentrations were broadly based on source–receptor models that are widely used for source apportionment. More specifically, the investigators based their work on a "spatial econometric" approach, incorporating a statistical accounting of emissions in the manner of economic analysis, adapted for the current purposes of associating emissions and air pollution levels.

APPROACH

The investigators' specific aims for the study were as follows:

1. To assess what portion, if any, of the observed reductions in ambient concentrations of $PM_{2.5}$ that occurred in the United States in the years 1999–2005 could be credited to emissions reductions resulting from the implementation of Title IV Phase 2 of the 1990 Clean Air Act Amendments; and

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. Richard D. Morgenstern of Resources for the Future, Washington, D.C., and colleagues. Research Report 168 contains both the detailed Investigators' Report and a Commentary on the study prepared by the Institute's Health Review Committee.

2. To develop a statistical modeling approach to link observed changes in emissions of SO_2 and NO_x from power plants to changes in $PM_{2.5}$ concentrations.

The models for the study were built using three datasets. Two of these, the EPA's Clean Air Markets database and the National Emissions Inventory database, provided inventories of source emissions of SO_2 and NO_x. The third, modeled as receptor data, consisted of air quality monitoring data from the EPA's Air Quality System.

For emissions, the Clean Air Markets data were classified by location and date and adjusted for any important National Emissions Inventory database variables (e.g., industrial emissions sources that opted in to the Acid Rain Program during the study period). The Air Quality System dataset was also sorted by date and location, and monthly average values were calculated.

Based on the location of the 193 Air Quality System monitors used to build the model, the investigators defined circular zones of radius up to 500 miles for tabulating emissions. The monthly average $PM_{2.5}$ readings at the monitors, the emissions in the zones surrounding the monitors, meteorologic variables from the Air Quality System database, and a large array of dummy variables and interaction variables (referred to as fixed-effects variables) to account for unmeasured emissions and factors influencing $PM_{2.5}$ concentrations were combined in the statistical modeling framework relating emissions of SO_2 and NO_x to measured $PM_{2.5}$.

Morgenstern and his team used a linear regression model to explore the statistical relationships between source emissions of SO_2 and NO_x and the monitored concentrations of $PM_{2.5}$. In these models, monthly average monitored concentrations of $PM_{2.5}$ were the "outcome," or "dependent," variable, modeled as a function of power plant emissions of SO_2 and NO_x in the circular zones and of the fixed-effects variables. The investigators first constructed a relatively simple model in which monthly average $PM_{2.5}$ measured at the monitors was modeled as a function of monthly power plant source emissions in the circular zones, monitor-site temperature (measured as deviation from mean temperature for the study period), and fixed-effects variables for monitor site, year, and month. Because the emissions of SO_2 and NO_x were highly correlated, the investigators chose to drop NO_x from their models.

The investigators then built more complex models that included interaction terms combining such factors as temperature and emissions or site and year, exploring the effects of zone rings of various sizes around the monitors, of the inclusion of fixed-effects variables for season, and so on. Their preferred model contained the main SO_2 emission variable, an interaction variable for temperature and emissions, circular emission zones up to 400 miles from the $PM_{2.5}$ monitors, a set of dummy variables for the effect of calendar month, and statistical adjustments for known violations of linear regression assumptions.

In addition to the use of mean squared error (MSE) as a measure of model fit, Morgenstern and colleagues also performed various external comparisons of their models. As an initial check, they applied the model to data for the year 2006, which had not been included in the data used to build the model. The investigators also compared the performance of the source–receptor model with the results of an EPA Regulatory Impact Analysis that used the CMAQ model to predict changes in ambient $PM_{2.5}$ based on reductions in emissions expected from implementation of the Clean Air Interstate Rule.

In order to evaluate the impact of the emissions reductions that occurred under Title IV of the Clean Air Act, the investigators simulated a counterfactual situation in which no mandated reductions in SO_2 occurred over the time period of the study. This simulation assumed that electric power plants covered under the Acid Rain Program continued to emit SO_2 at the same rate as before the regulations and that consumption of electric power was the same as the actual consumption for the time period. They then ran their preferred model, built from the source–receptor analysis of actual 1999–2000 SO_2 emissions and ambient $PM_{2.5}$ data, with the counterfactual SO_2 emissions data and compared the results with the actual measurements.

RESULTS

The investigators' preferred model, which included a temperature–emissions interaction term and monthly dummy variables, predicted monthly average $PM_{2.5}$ concentrations at the Air Quality System monitoring locations with reasonable precision (MSE = 9.37 for emission zones within a 400-mile radius, indicating a mean prediction error at a "typical" monitor of $\sqrt{9.37} = \pm 3.06$ $\mu g/m^3$). The model performed even more favorably when applied

to actual data for the 164 monitors in the 2006 dataset that corresponded to monitors in the 1995–2005 dataset, producing more precise predictions (MSE = 7.85 for 2006 versus 9.37 for the 1999–2005 data). When the model was applied to 2006 data for the remaining 445 monitors that did not provide data for the original model-building exercise, prediction improved even further (MSE = 7.19). By comparing their counterfactual scenario with the actual regulatory scenario, the investigators calculated that Title IV of the Clean Air Act resulted in an estimated reduction in ambient $PM_{2.5}$ concentrations (averaged across the eastern United States) of 1.07 $\mu g/m^3$ between 1999 and 2005 (or 0.89 $\mu g/m^3$ on a population-weighted basis).

CONCLUSIONS

In its independent review of the study, the HEI Health Review Committee found that the primary strength of the study was that it was data-driven and observational, rather than simply using complex modeling techniques to assess the impact of a regulatory intervention. The Committee believed that the authors were appropriately careful about the inferences they drew from their work and reasonably cautious about their findings. It also found that Morgenstern and his team made a good-faith effort to address the general scientific questions about the effects of regulations on air quality that they had set out to address.

The Committee expressed a number of concerns about the development of the models and their potential application to datasets other than those used to build them. Although the Committee felt there was some value to the investigators' overall approach, the Committee found it difficult to fully assess the potential application of these models to air quality management or impact assessment.

Despite these limitations, it was the Committee's judgment that the investigators' work contributed to the discussion of what portions of PM reductions can be attributed to an emissions reduction program, with an approach that might be a useful alternative to atmospheric models in some applications. The Committee also noted that a model that estimates ambient air quality changes secondary to emissions changes might also be applied to estimating changes in criteria pollutants secondary to regulations aimed at reducing industrial emissions or even greenhouse gas emissions. This research and these models might also be able to provide useful information to organizations that want a quick estimate of how much specific emitting facilities affect specific pollutant monitors or even communities, although atmospheric pollution dispersion models might be more readily applied.

INVESTIGATORS' REPORT

Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments

Richard D. Morgenstern, Winston Harrington, Jhih-Shyang Shih, and Michelle L. Bell

Resources for the Future, Washington, D.C. (R.D.M., W.H., J.-S.S.); Yale University, New Haven, Connecticut (M.L.B.)

ABSTRACT

In this study, we sought to assess what portion, if any, of the reductions in ambient concentrations of particulate matter (PM*) ≤ 2.5 μm in aerodynamic diameter ($PM_{2.5}$) that occurred in the United States between the years 1999 and 2006 can be attributed to reductions in emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x) resulting from implementation of Phase 2 of Title IV of the 1990 Clean Air Act Amendments. To this end, a detailed statistical model linking sources and monitors over time and space was used to estimate associations between the observed emissions reductions and improvements in air quality.

Overall, it turned out to be quite feasible to use relatively transparent statistical methods to assess these outcomes of the Phase 2 program, which was designed to reduce long-range transport of emissions. Associations between changes in emissions from individual power plants and monitor-specific estimates of changes in concentrations of $PM_{2.5}$, our indicator pollutant, were highly significant and were mostly of the expected relative magnitudes with respect to distances and directions from sources. Originally estimated on monthly data for a set of 193 monitors between 1999 and 2005, our preferred model performed equally well using data for the same 193 monitors for 2006 as well as for an additional 217 monitors not in the original set in 2006.

Although substantial model uncertainty was observed, we were able to estimate that the Title IV Phase 2 emissions reduction program implemented between 1999 and 2005 reduced $PM_{2.5}$ concentrations in the eastern United States by an average of 1.07 $\mu g/m^3$ (standard deviation [SD] = 0.11 $\mu g/m^3$) compared with a counterfactual case defined as there having been no change in emission rates per unit of energy input (1 million British thermal units [BTUs]). On a population-weighted basis, the comparable reduction in $PM_{2.5}$ was 0.89 $\mu g/m^3$.

Compared with the air quality fate and transport models used by the U.S. Environmental Protection Agency (EPA) to estimate air quality improvements associated with the Clean Air Interstate Rule (CAIR) for 2010 and 2015, when baseline $PM_{2.5}$ concentrations were expected to be about one-third lower, our statistical model yielded roughly similar results per ton of SO_2 reduced, well within the estimated confidence intervals of the models.

We have proposed a number of steps to advance air quality outcomes research using statistical methods. Specifically, we have emphasized the value of updating our analysis with post-2005 data to try to corroborate our findings. We have also recommended extending the work on air quality outcomes to include changes in health outcomes that might be associated with the implementation of Title IV Phase 2.

INTRODUCTION

To understand the true impact of major air pollution policies, it is critical to examine the outcomes of these policies using evidence-based assessments. Much of the existing literature has been based on intervention studies involving short-term actions at local scales, such as the Beijing (2008) or Atlanta (2002) Olympics, where relatively large, short-term changes in ambient pollution

This Investigators' Report is one part of Health Effects Institute Research Report 168, which also includes a Commentary by the Health Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. Richard D. Morgenstern, Resources for the Future, 1616 P St. N.W., Washington, DC 20036; e-mail: *morgenstern@rff.org.*

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83234701 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.

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

concentrations facilitated detection of a statistical signal linking changes in air quality and health outcomes. Detection of such a statistical signal is more challenging in the case of long-term, regional emission controls on sources located quite far from the areas they were intended to benefit. That is the subject of this research.

In 1990, as part of a broad series of changes, the Clean Air Act was amended to establish a national program to control acid rain precursors. Title IV of the 1990 Amendments called for reducing emissions of SO_2 and NO_x from power plants. Although SO_2 and NO_x are emitted from many sources, power plants are responsible for emitting the vast majority of the SO_2 and a significant percentage of the NO_x found in the atmosphere.

Title IV aimed to reduce SO_2 emissions through a two-phase program. After Phase 1 emission reduction requirements had become effective for a subset of plants, the "big, dirty" units, Phase 2 tightened annual emission limits on the Phase 1 units and set new limits for more than 2,000 smaller, cleaner units in all 48 contiguous states and the District of Columbia. The Phase 2 requirements, which became effective in 2000, represented the largest reduction in criteria pollutants to occur over a short period in the history of clean air regulation up to that time.

The Phase 2 reductions occurred at a time when an advanced, spatially detailed monitoring system was in place to detect respirable fine particles ($PM_{2.5}$), thus creating a unique opportunity to develop a detailed outcomes analysis focusing on long-range transport of SO_2 and its conversion into $PM_{2.5}$. The current study covers the seven-year period beginning in 1999, one year before the formal start of phase 2, and extending through 2005.

In air pollution outcomes research, the key links in the accountability chain are those between regulatory action and emissions, emissions and air quality, air quality and exposure, and exposure and human health outcomes (Committee on Research Priorities for Airborne Particulate Matter 1998; Health Accountability Working Group 2003). Although the first two links are routinely used by the EPA in individual rulemakings and evaluative studies, the analyses underlying them are generally "ex ante" in nature (i.e., predicting the effects of proposed regulations on ambient air quality). The third and fourth links have the strongest empirical support (Samet et al. 2000).

By its very nature, however, an outcomes assessment demands an "ex post" examination (i.e., looking after the fact at the actual observed effects of enacted regulations on ambient air quality). Further, to link emissions reductions to air quality improvements at distant locations, one needs to develop a spatially disaggregated counterfactual analysis, that is, an analysis of what the pollutant concentrations would have been at various locations in the absence of the policy that was hypothesized to have caused the air quality improvements. In the case of the current study, such an analysis also had to account for local or other policy changes beyond those associated with Title IV. Once a reliable counterfactual is developed, it becomes possible to compare predicted with observed pollutant concentrations across a study area.

A series of regulatory actions begun by the EPA in 1993 established the implementation program for Title IV and the data collection requirements needed to manage the program (as well as the framework for the current study):

- The Acid Rain Core Rule (1993) defined two phases: Phase 1 (1995) required reductions from the "big, dirty" power plants, and Phase 2 (2000) required reductions from all other power plants affected by Title IV. Continuous emissions monitors for SO_2 and NO_x were mandated for virtually all regulated stacks, covering more than 90% of total SO_2 emissions and about 60% of total NO_x emissions in the eastern United States.
- The 1997 ambient air quality standard for $PM_{2.5}$ specified monitoring requirements for this new criteria pollutant. Although the standard itself was delayed in litigation until 2001, ambient monitoring requirements became effective in 1999. By 2004, there were 250 $PM_{2.5}$ monitoring sites in the EPA's Air Quality System.
- CAIR, promulgated in 2005, mandated an additional 45% reduction in SO_2 emissions by 2010 and further reductions in later years. Even though CAIR was subsequently overturned by the courts, the banking provisions of the rule, which allowed utilities to use current excess emission reductions in future years, when it was expected to be more costly to make such reductions, further increased the incentive to reduce SO_2 emissions in 2005 and beyond.

During the 1999–2005 period, other important emissions reduction programs were also in place, principally those designed to reduce NO_x emissions from mobile and area sources. However, the power plant requirements had quite dramatic and identifiable impacts on total emissions and, we expected, on concentrations of $PM_{2.5}$, our indicator pollutant. The coincidental timing of the implementation of these policies to regulate emissions of SO_2 and NO_x as well as concentration of $PM_{2.5}$, coupled with the enhanced, spatially detailed data available from the $PM_{2.5}$ monitoring network, presented a unique opportunity to conduct a rigorous outcomes analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments.

SPECIFIC AIMS

The overall goal of this study was to assess what portion, if any, of the observed reductions in ambient concentrations of $PM_{2.5}$ that occurred in the United States in the years 1999 through 2005 could be credited to emissions reductions resulting from the implementation of Title IV Phase 2 of the 1990 Clean Air Act Amendments.

Average ambient concentrations of $PM_{2.5}$ in the United States declined by 10% between 1999 and 2002. In the southeast United States, average ambient concentrations of $PM_{2.5}$ declined by about 20%. Over the same period, annual power plant emissions of SO_2 declined by 2 million tons, more than 15%. As a result of rising natural gas prices, a series of nuclear-plant maintenance issues, and mandated shutdowns to install NO_x controls in certain facilities, coal use increased somewhat in 2003, with a concomitant increase in SO_2 emissions in that year. The announcement of CAIR had its initial impact in early 2004, when SO_2 emissions again started to decline. Power plant NO_x emissions also declined over this period, albeit in a largely monotonic pattern, representing approximately a 15% reduction between 1999 and 2005.

The current study has attempted to account for these occurrences by establishing a causal relationship between the emission reductions and the improvements in air quality using statistical techniques to link sources and monitors. Specifically, we developed a statistical modeling approach to link observed changes in SO_2 and NO_x emissions to changes in concentrations of $PM_{2.5}$.

STATISTICAL MODELING APPROACH TO DEMONSTRATING LINKAGE BETWEEN SOURCES AND MONITORS

In this section, we describe the statistical method we used to establish a causal relationship between the emissions reductions associated with Title IV Phase 2 and the observed changes in $PM_{2.5}$ concentrations. The section also presents background information, including a brief discussion of the relevant literature; a description of the data used in the study, including the sample selection and averaging periods; a description of the methodology used; the results obtained; the results of a test of our modeled relationship for 2006 (an out-of-sample year beyond our basic dataset); the development of various counterfactual comparisons; and, finally, the study's conclusions.

BACKGROUND

There are several points in the logical pathway from regulation to human health response that can be investigated in outcomes research (also known as accountability research). The Health Effects Institute used a National Research Council framework to evaluate associations between air pollution and health outcomes to identify links in the framework for which outcomes can be studied (Health Accountability Working Group 2003). Regulatory action can be evaluated to determine if emissions controls have in fact been established. Emissions can be measured to determine if emissions of air pollutants and precursors of air pollutants have decreased (or increased). Ambient air quality can be measured to determine if changes in sources and emissions have resulted in changes in concentrations of air pollutants. Personal exposures can be measured to determine if exposures to air pollutants have changed (which could be a result of changes in ambient concentrations or changes in behaviors, such as staying indoors, that have altered exposure patterns). Changes in exposures can affect dose-to-target tissues, which can result in turn in changes in human health responses (Health Accountability Working Group 2003). Outcomes research can study these links (regulatory action $\rightarrow$ emissions $\rightarrow$ ambient air quality $\rightarrow$ personal exposure $\rightarrow$ dose-to-target tissues $\rightarrow$ human health response) individually or in various combinations. Some studies also study the economic impacts of regulatory actions and health responses.

In addition to the links in the accountability chain, the study of accountability has several other aspects. Analysis can consider short-term impacts, such as an immediate impact after a specific occurrence. This approach is sometimes referred to as an intervention study. Analysis can also consider longer-term impacts, such as the impact of a regulatory action over several years. Spatial scales can be local (e.g., city-level) or regional–national (e.g., the United States). The pollutants and emissions considered are often specific to the spatial area of study. As an example, researchers could investigate whether local changes in transportation in a given community have affected air quality and health response within the community. To date, few studies have examined the long-range transport of pollutants. In most cases, studies investigate the link between a regulatory action and the changes in air quality that were hypothesized (generally because of the date the action went into effect) to be a result of the action rather than estimating how a regulatory action has altered emissions and then pollutant concentrations. Four types of accountability studies are summarized below.

Intervention Studies: Short-Term Action at a Local Scale

The most common approach to accountability is to study short-term changes in air pollution and subsequent health effects resulting from a substantial change in air quality control. These studies take advantage of the unique

opportunities presented by interventions to examine localized health effects before and after regulatory action.

Several researchers have investigated health outcomes surrounding Olympic Games, for example, which often coincide with dramatic improvements in local air quality.

Heart rate variability, a marker of cardiac autonomic function, and personally monitored $PM_{2.5}$ concentrations were measured for 11 taxi drivers (ages 27 to 41 years) during a 12-hour work period recorded between 9 am and 11 pm before (May 26 to June 19), during (August 11 to September 5), and after (October 27 to November 14) the 2008 Summer Olympic Games in Beijing (Wu et al. 2010). This population was of particular interest given its high exposure to air pollution. Significant air quality control measures were implemented for the Games, including restrictions on industrial emissions and motor vehicle use. Three measures of heart rate variability were estimated: (1) SD of normal-to-normal intervals, (2) low-frequency power (0.04–0.15 hertz, equaling 1 cycle per second), and (3) high-frequency power (0.15–0.40 hertz). All three measures were significantly higher during the Olympics compared with either baseline period. Pollutant concentrations were lower during the Games than during the baseline periods for personal $PM_{2.5}$ exposures and ambient measures of $PM_{2.5}$, carbon monoxide, nitrogen dioxide (NO_2), and nitric oxide. Heart rate variability indicators were significantly associated with personal $PM_{2.5}$ exposures.

Another study of the Beijing Olympics found that the number of asthma outpatient visits for adults was lower during the Games (August 8 to September 29, 2008) than during a baseline period more than a month earlier (June 1 to 30) and a baseline period immediately before the Games (July 1 to August 7) when some of the vehicle use restrictions, but not the full industrial restrictions, had been put in place (Li et al. 2009).

Friedman and colleagues (2001) evaluated whether vehicle use policies temporarily enacted for the 1996 Summer Olympic Games (July 19 to August 4) in Atlanta affected acute care visits and hospitalizations for residents 1 to 16 years of age. Baseline conditions were assessed for the 4 weeks before and 4 weeks after the Games. The number of acute care visits was lower during the Games than during the baseline periods for asthma and non-asthma causes; the decrease in asthma-related visits was statistically significant. Ozone (O_3) concentrations and traffic counts also decreased during the Games compared with those of the baseline periods.

The 2002 Asian Games, held in Busan, Korea, provided another opportunity to investigate how air pollution concentrations changed in response to temporary restrictions on industrial air pollutants and vehicle use and the subsequent effects on human health (Lee et al. 2007). The Games period was compared with the 3-week period before (September 8 to 28) and the 3 weeks after (October 15 to November 4) the Games. Time–series analysis was used to estimate the associations between air pollutants — including NO_2, O_3, carbon monoxide, SO_2, and PM $\leq$ 10 µm in aerodynamic diameter (PM_{10}) — and hospital admissions for asthma for persons less than 15 years old; the analysis included interaction terms to allow effect estimates to vary by time period. Concentrations for almost all pollutants were 1% to 25% lower during the Games than during the baseline periods but rose again during the same dates in the following year. Similarly, hospitalizations of children for asthma were lower after the Games but increased thereafter. Associations with a risk of children's hospitalization were observed for several pollutants (PM_{10}, SO_2, NO_2, and O_3).

In Dublin, particulate concentrations in the form of black smoke and mortality were assessed for the 72 months before and 72 months after September 1, 1990, the date the Irish government imposed a ban on bituminous coal (Clancy et al. 2002). Pollutant concentrations were lower after the ban, as were mortality rates. The mortality analysis was adjusted for respiratory epidemics, weather, and temporal changes in the age distribution. Non-trauma mortality rates were 5.7% lower after the ban compared with those of the period before, cardiovascular mortality was 10.3% lower, and respiratory mortality was 15.5% lower. Reductions in mortality rates were observed in multiple age groups (< 60, 60–74, and > 75 years old).

Studies of National Policies: Long-Term Action at a National Scale

The EPA estimated the costs of the Clean Air Act of 1970 over a 20-year period (from 1970 to 1990) at $0.52 trillion and the benefits at $5.6 trillion to $49.4 trillion (central estimate $22.2 trillion) (U.S. EPA 1997). The study design used an array of modeling systems, including macroeconomic, emissions, and air quality modeling. Health benefits were estimated for a variety of causes, including hospital admissions, restricted activity, work loss days, and hypertension; monetary benefits were dominated by averted mortality. Benefits other than human health were also examined, such as effects on visibility and agricultural yield. The report acknowledges that not all consequences from air pollution were quantified, because of the state of scientific knowledge at the time. These consequences included changes in pulmonary function from O_3, materials damage from NO_x, reproductive effects from lead exposure, and air toxics effects. Analysis isolated the impacts of the Act specifically rather than of air quality

control policies more generally. The study's baseline scenario assumed continuation of the emission rates in place prior to enactment of the Clean Air Act of 1970. A follow-up study examined the prospective benefits and costs of the Clean Air Act from 1990 to 2010 (U.S. EPA 1999). Air quality modeling was used to estimate future pollution concentrations. The benefits of the Act were estimated at $110 trillion in 2010 (in $1990), and the direct costs were estimated at $27 trillion, for a benefit–cost ratio of 4 to 1. (For further analysis of the two EPA studies, see Krupnick and Morgenstern [2002].)

A study by Chay and colleagues (2003) used econometric techniques to examine whether changes in the degree of attainment of the National Ambient Air Quality Standards for total suspended particulates for 1971–1972 were associated with changes in adult mortality rates over the same period. Changes in total suspended particulate concentrations were observed; changes in adult mortality were not. The authors suggested that the results should be viewed with caution because of limitations in the study design. In other work, they found that reductions in total suspended particulates corresponding to the time period of the 1981–1982 U.S. recession were associated with lower infant mortality rates, especially for neonatal births (Chay and Greenstone 2003).

A study by Lin (2010) is of particular interest, given its similarities to the current study. In the context of a case study of a regional NO_x cap and trade program, Lin used a spatial econometric approach to address long-range transport (rather than the more traditional use of regional air quality modeling) in order to estimate how O_3 precursor emissions in one location can affect externalities on O_3 in another state. Data included hourly O_3 measurements for July 1990; daily maximum temperatures; county-level emissions of NO_x and volatile organic compounds (the precursors to O_3) for 1990, 1996–1999, and 2001; and county-level population and per capita personal income for 1969–2001. Overall, the results suggested that a NO_x cap and trade program without spatial restrictions is not an effective policy mechanism to decrease O_3 concentrations.

Lin (2010) highlighted several of the challenges in applying atmospheric models to address changes in pollutant concentrations, particularly for a pollutant with complex chemistry such as O_3. These included the absence of estimates of statistical uncertainties, limited ability to capture stochastic elements, limited accuracy (such as with boundary conditions), and potential problems in supporting input data and diagnostic methods to evaluate model performance. The application of econometrics rather than air quality modeling obviated the need to make assumptions about the functional form of the relationships between O_3 precursors and resulting concentrations, such as those needed in atmospheric chemistry models.

Potential Policy Action: Long-Term Action at a Local or National Scale

Several studies have estimated the health benefits that would occur under hypothetical air pollution control policies. Below we summarize three such studies comparing "business-as-usual" scenarios with various potential policies. The first two examined compliance with air quality guidelines or regulations as the control, and the third assumed a control policy to lower O_3 and PM_{10} concentrations. The studies were local (Hong Kong and three Latin American cities) and national (the United States) in scope and considered long time frames.

Hedley and colleagues (2008) estimated air pollution quality in Hong Kong and compared the estimates with World Health Organization guidelines. The study used a unique approach of estimating air quality through visibility from photographs. Estimates from existing local studies using time-series models were applied to link the estimated changes in concentrations of NO_2, SO_2, PM_{10}, and O_3 to the risk of various human health outcomes. Pollutant concentrations were adjusted to account for the contribution of NO_2 and SO_2 to PM_{10}. Baseline health risks were assessed through government and hospital data. The authors estimated all-cause mortality, respiratory and cardiovascular hospital admissions, and family doctor visits for respiratory and upper respiratory tract infection. The economic costs of these health outcomes were also estimated. Direct costs were calculated for illness for public and private hospital admissions, public outpatient consultations, and family doctor visits. Productivity losses were estimated for loss of life and lost work time for adults (ages 15–64 years) using labor-force and employment rates and gender-adjusted salaries. Sensitivity analyses were conducted with respect to a multi-pollutant adjustment for particles. Results indicated that if air pollution concentrations in Hong Kong were reduced to meet World Health Organization guidelines, health benefits would include an averted 1,335 deaths, 60,587 hospital bed days, and 6.7 million doctor visits for respiratory symptoms on an annual basis, saving $240 million (U.S.) annually. The study recognized limitations, such as certain health outcomes (e.g., detriment to lung function growth and sudden infant death syndrome) and medical practices (e.g., self-medication) that were not included.

Hubbell and colleagues (2006) investigated the health benefits that would have occurred from 2000 to 2002 in the United States had the measurements from all ambient monitors complied with the National Ambient Air Quality

Standards for O_3. The authors used existing associations between pollution and health from epidemiologic studies and existing evidence of the value of averted health outcomes from economic studies. Over the 3 years of the study, an estimated 800 deaths, 4,500 hospitalizations and emergency department admissions, 900,000 school absences, and more than a million minor restricted activity days would have been avoided through compliance with the O_3 standards. The estimated economic impact of these health outcomes was $4.9 billion to $5.7 billion.

Annual O_3 and PM_{10} concentrations were estimated over a 20-year period (2000 to 2020) for three large Latin American cities — Santiago, Chile; Mexico City, Mexico; and São Paulo, Brazil (Bell et al. 2006). Two policy scenarios were considered: (1) a business-as-usual scenario based on emissions and regulatory trends current at the time of the study and (2) a modest air control policy. Differences in emissions concentrations across time for each city for the two policy scenarios were compared with epidemiologic evidence of how O_3 and PM_{10} concentrations are associated with human health. Local human health studies were applied when possible. It was estimated that the air pollution control policy would avert more than 156,000 deaths, 4 million asthma exacerbations, 300,000 children's medical visits, and almost 48,000 cases of chronic bronchitis over the 20-year period across the three cities. Two approaches, willingness-to-pay and cost-of-illness, were used to estimate the economic impact of these health outcomes at $21 billion to $165 billion.

Changes in Risk Estimates of Pollution and Health over Time: Long-Term Action at a Local or National Scale

Whereas most accountability studies examine how changes in air pollutant concentrations affect overall human health outcomes (e.g., lower PM_{10} concentrations that result in lower mortality rates), some studies examine how the association between a given change in air pollutant concentration and health outcomes changes over time. Such research investigates whether a given increase in a given pollutant is more (or less) harmful for some time periods than for others, which is a separate issue than whether lowered overall pollutant concentrations result in health benefits. Because of air pollution control polices, the nature of air pollution mixtures can change over time, possibly affecting their toxicity. This issue is of particular importance for PM, because the chemical structure of PM varies spatially and temporally. Two examples of this type of study were a local study of the city of Erfurt, Germany, and a national study of the United States. Both estimated temporal trends in PM's association with mortality and air control policies.

Breitner and colleagues (2009) examined the associations between daily mortality risk and PM of various size distributions (PM_{10}, $PM_{2.5}$, and ultrafine PM) and how the risk per unit pollutant concentration changed over time during a period in which air quality improved in Erfurt. Statistical models were used to relate air pollution concentrations to daily mortality risk for two time frames: (1) 1991 to 2002 for NO_2, carbon monoxide, and PM_{10} and (2) 1995 to 2002 for ultrafine PM and $PM_{2.5}$. After developing models of how day-to-day variations in mortality risk were associated with day-to-day variations in pollutant concentrations, additional models were fitted, allowing the effect estimates to vary temporally with indicator variables for various time periods (October 1991 to August 1995, September 1995 to February 1998, and March 1998 to March 2002). Associations were observed between mortality risk and ultrafine PM but not PM_{10} or $PM_{2.5}$. The observed associations per unit pollutant concentration were generally higher in the middle time period, with lower estimates in the later period (data were not available for all pollutants in the earlier time period). The estimated changes in health risks coincided with broad reforms in air pollution programs associated with the reunification of Germany, which changed the distribution of pollutant sources and the characteristics of particles and generally lowered air pollution concentrations.

Another study that examined changes in relative risk estimates, as opposed to changes in overall health outcomes, evaluated the association between PM_{10} and mortality risk over a 14-year period (1987 to 2000) using data from the National Morbidity, Mortality, and Air Pollution Study (Dominici et al. 2007). Estimates were generated separately for three time periods — 1987 to 1994, 1995 to 2000, and 1987 to 2000 — for 96 to 100 U.S. urban communities. Results provided suggestive evidence that the relative risk rates of mortality associated with fixed increments of PM_{10} concentration have declined over time in the eastern United States.

Comparisons with the Current Study

Whereas many outcomes studies (such as the studies of the Olympic Games described above) explored localized events and short time frames, our work examined a relatively large spatial scale (the eastern United States) and long time frame (the seven years from 1999 through 2005) in relation to Title IV Phase 2 of the 1990 Clean Air Act Amendments. A distinctive feature of our work was the emphasis on the emissions link in the accountability chain (regulatory action → emissions → ambient air quality → personal exposure → dose to target tissues → human health response), specifically the link between SO_2 and NO_x

emissions and $PM_{2.5}$ concentrations. The emissions–air quality link was explicitly modeled by way of statistical methods. Our study is most similar to the studies of national policies described above, because it addressed a specific regulatory action over a large spatial scale and a long time frame.

At the same time, our approach involved statistical modeling as opposed to the air quality modeling that was used in, for example, the EPA study on the costs and benefits of the Clean Air Act (U.S. EPA 1999). Both methods have strengths and limitations. Although the knowledge base varies by pollutant, air quality modeling generally benefits from the incorporation of scientific understanding of the physical and chemical transformations of pollutants and their atmospheric dispersion. Tropospheric O_3 concentrations, for example, have generally been better estimated by air quality modeling than have the various chemical components of $PM_{2.5}$. Because statistical modeling is not explicitly based on the physical and chemical transformations of pollutants or their atmospheric dispersion, the method does not require a full understanding of these complex processes (which include chemistry that is often nonlinear). However, this same advantage can be problematic if the modeling fails to fully address issues that result in different relationships among the variables (e.g., meteorology, emissions, and topography) for locations and time periods other than those of the data used to develop the models. Clearly, consistent findings for outcomes studies based on both modeling methods would provide stronger evidence than findings from either method alone.

DATA

This section describes the emissions and ambient data used in this study.

Emissions Data

The SO_2 and NO_x datasets we used were from the EPA's Clean Air Markets (CAM) Division. The CAM datasets include information on unit level (point source) emissions for all electric utilities participating in the EPA's Acid Rain Program or the NO_x budget trading program as well as industrial sources that opted into the Acid Rain Program. Monthly data represent aggregations of sub-hourly readings collected by continuous emissions monitors, which are required of all utility emission sources in the Acid Rain Program. We also obtained the National Emissions Inventory (NEI) for 1999 and 2002. The NEI contains estimated annual emissions for all point sources, mobile sources, and area sources in the United States. We used this information to crosscheck the CAM datasets. In addition to the point source SO_2 and NO_x emissions for utilities participating in the Acid Rain Program, we also examined the point sources not participating in the Acid Rain Program and not appearing in the CAM database. For mobile and area sources, we considered using NEI county-level estimates of mobile and area sources scaled by estimated changes in population and vehicle stock. However, because these data were only available on an annual basis, they were not suitable for our modeling framework, which was based on monthly observations. As described in the Statistical Model section below, we used an extensive set of fixed-effects variables in our modeling in order to take these local sources into account.

Tables 1 and 2 show summary emissions data arranged in two ways. Table 1 shows mean, maximum, and minimum total monthly SO_2 and NO_x emissions for electric generating stations. Table 2 shows SO_2 emissions data by emission ring around the 193 monitors in our final dataset.

Table 1. Characteristics of Monthly Emission Dataset: Electric Generating Stations[a]

	Emissions (Tons/Month)		
	Mean	Maximum	Minimum
SO_2	656	25,556	0
NO_x	256	11,150	0

[a] Number of generating stations = 1314; number of generating stations with SO_2 emissions = 1022.

Table 2. Characteristics of Monthly Emission Dataset: Emission Zones[a]

	Average SO_2 Emissions in Each Zone (Tons/Month)			
Ring	Mean	Zeros (%)	Mean Nonzero	Maximum
0–12.5 miles	176	88	1,523	25,556
12.5–25 miles	274	86	1,942	25,556
25–50 miles	831	70	2,786	33,773
50–100 miles	2,639	37	4,200	48,641
100–150 miles	3,915	27	5,409	71,500
150–200 miles	4,286	24	5,609	80,999
200–300 miles	12,717	14	14,873	145,731
300–400 miles	15,982	20	20,012	197,157

[a] Number of zones = up to 64 per monitor, depending on specification.

(Instead of showing data for all 64 emission zones, we grouped them by ring.) The first column of Table 2 shows mean monthly emissions by ring; as the rings got larger, the average emissions increased rapidly. The second column shows the percentage of zones with zero emissions by ring; overall, out of more than 1 million data points (from 193 monitors and 64 emission zones over 84 months), about half (64%) had zero emissions. (Not surprisingly, the small area of the closer-in zones ensured that the percentage of zeros was high.) The third column shows the mean emissions in the nonzero zones by ring; these also increased as the rings got larger, though not as fast as the unconditional means. The fourth column shows the maximum emissions for any zone at any time by ring.

Opt-In Emissions Industrial sources of SO_2 were allowed to opt into the Acid Rain Program. This fact complicated our effort to combine emissions in the NEI and CAM databases. Emission readings from point sources not participating in the Acid Rain Program were available on an annual basis in the NEI. If significant numbers of new industrial SO_2 sources were found to be opting into the Acid Rain Program, we would have had to account for their emissions in the NEI database; otherwise they could have been double counted.

Table 3 shows SO_2 and NO_x emissions from "old" sources (i.e., those that were part of the Acid Rain Program before January 1, 1999) and "new" sources (i.e., those that were not part of the program until January 1, 1999, or later). As can be seen, almost no SO_2 emissions and only a very small amount of NO_x emissions came from new sources opting into the program during the time period we studied (1999–2005). The percentage data in the table represent an upper bound on the percentage of emissions that could have been double counted. Furthermore, of the NO_x emissions from sources that opted in on or after January 1, 1999, 47% were from electric utility sources, which means they were actually new sources just beginning operations. As shown in the last column, the NO_x emissions from new non-electric utility sources opting into the program were less than 2.5% of emissions from existing utility sources. Even this estimate overstates the problem. CAM classified sources as electric plants or not electric plants using the U.S. government's Standard Industrial Classification (SIC) code system. In the CAM database, the SIC code was missing for a substantial number of entries. To be conservative, then, all the missing codes were assigned to the non-utility sector. We therefore concluded that the potential double counting of opt-in emissions could be ignored.

Other Point Sources of SO_2 and NO_x To obtain estimates of other emission sources, we turned to the NEI, which provides estimates every third year of annual emissions of all pollutants from all sources. From the point-source inventories for 1999 and 2002, we compared SO_2 and NO_x emissions county by county to determine how much change had taken place. The results are shown in Table 4. Comparing Tables 3 and 4, it appears that non-CAM emissions were about 20% of CAM emissions for SO_2 and about 50% for NO_x. Between 1999 and 2002, total annual non-CAM emissions of SO_2 and NO_x each declined by about 10%. Thus, the declines in non-CAM emissions of SO_2 and NO_x were 12% and 27%, respectively, of the declines in CAM emissions. Overall, the changes in CAM emissions accounted for the overwhelming bulk of

Table 3. Opt-In Emissions of SO_2 and NO_x, 1999–2005

	SO_2 Emissions (ktpy)[a]			NO_x Emissions (ktpy)[a]			
					New Plants		
Year	Old Plants	New Plants	New/Old (%)	Old Plants	Electric	Nonelectric	New Nonelectric/ Old (%)
1999	11,600	0.0	0.00	4,814	20	26	0.54
2000	10,500	1.2	0.01	4,420	31	29	0.66
2001	9,890	4.7	0.05	4,021	37	28	0.70
2002	9,540	14.2	0.15	3,810	50	31	0.81
2003	9,958	16.5	0.17	3,498	52	80	2.29
2004	9,636	15.2	0.16	3,098	56	76	2.45
2005	9,640	12.9	0.13	2,964	59	72	2.43

[a] ktpy indicates kilotons per year.

Table 4. Non-CAM SO_2 and NO_x Emissions (in Kilotons) from Point Sources, 1999–2002

	SO_2	NO_x
Total emissions, 1999	2431	2438
Total emissions, 2002	2185	2168
Change in emissions, 1999–2002	246	270
Sum of absolute county-level changes	983	919

aggregate emission changes in SO_2 and NO_x between 1999 and 2002. (Within the non-CAM totals, however, there was apparently a substantial redistribution of emissions among counties. When the absolute differences in emissions of SO_2 and NO_x between 1999 and 2002 were summed across counties, shifts were found of almost a million tons of each. To account for these shifts in local emissions, we adopted a fixed-effects model, as described in the Model Results section below.)

Completion of Non-Point Source Inventory Nearly all emissions of SO_2 are from point sources. For NO_x, however, significant quantities also come from on-road and non-road mobile sources. The latter sources are difficult to monitor, and their quantities must be estimated based on experience with emissions measurements from relatively small samples of specimen equipment (such as motor vehicles and construction equipment), coupled with estimates of its use. As with the non-utility point sources discussed above, only annual estimates are available. Thus, as described in the Statistical Model section, below, we proxied non-utility emissions by way of an extensive set of fixed-effects variables in our modeling.

Ambient Data

We obtained daily $PM_{2.5}$ and temperature data (1999–2005) from datasets maintained in the EPA's Air Quality System. Among other items, the datasets contained information on monitor location (by state and county), hourly temperature, and, for $PM_{2.5}$, sample duration, unit of measurement, sampling method, date, and sampling frequency. At most monitors, $PM_{2.5}$ was measured every three days; at some, especially in urban areas, it was measured daily. At nearly every monitor, however, there were periods, often lengthy, when no $PM_{2.5}$ data were collected.

During the study period, the national $PM_{2.5}$ monitoring network was still being built. Table 5 shows the numbers of monitors by year and of months in the year with at least one $PM_{2.5}$ reading. Table 6 shows the numbers of monitors, by year, that had at least *n* daily readings in any month for various values of *n*. (Some monitors occasionally reported two readings in a single day. In those cases we took the mean of the two and treated it as a single reading.) We note especially a jump of 156 monitors with usable readings from 1999 to 2000. After 2000, the increase in the frequency of readings per month was more gradual, reaching a peak in 2002 and then plateauing or even declining slightly.

Table 5. Expansion of the $PM_{2.5}$ Monitoring Network in the EPA's Air Quality System

	Monitors by Months of Usable Data					
	1–3	4–6	7–9	10–11	12	Total
1999	57	115	105	134	548	959
2000	26	25	49	64	951	1115
2001	26	26	39	23	1022	1136
2002	18	36	20	12	1038	1124
2003	78	43	25	19	958	1123
2004	29	18	26	10	958	1041
2005	26	34	75	26	882	1043

In evaluating the data and choosing receptors and monthly data for further analysis, we initially considered three criteria:

- Frequency of observations per month
- Distribution (length and frequency) of gaps between successive daily observations
- Data completeness

Table 6. Increases in $PM_{2.5}$ Monitoring Frequency in the EPA's Air Quality System

	Minimum Number of Readings in Any Month				
	0–2	3–6	7–9	10+	Total
1999	318	401	158	82	959
2000	185	392	372	166	1115
2001	106	331	510	189	1136
2002	77	312	510	225	1124
2003	107	286	506	154	1123
2004	61	310	530	140	1041
2005	90	311	514	128	1043

By far the most important consideration for our selection of monitors was data completeness. Our ambient data consisted of a panel dataset with multiple time-series observations on the same monitors. We sought to develop a balanced panel (i.e., one with observations for the same periods for each monitor) for two reasons: (1) because we were undertaking to estimate a model that allowed error terms to be serially correlated, missing observations took on great importance, and (2) the exclusion of a monitor because of a missing monthly observation would create a potential for selection bias. Furthermore, inasmuch as most of the changes in emissions and air quality occurred in 1999, the first year of our study period, we aimed to focus on a sample for the entire 84-month period (1999 to 2005) rather than only the 72-month sample (2000 to 2005).

Unfortunately, surprisingly few monitors provided a complete 84-month sample. Table 7 shows the numbers of monitors, at various minimum monthly monitoring frequencies, available for our initial 84- and 72-month datasets. As can be seen, to have even a minimum monthly frequency of one observation in an 84-month sample, we would lose nearly 80 percent of the monitors in the network in 1999; even for a 72-month sample, we would still lose 60 percent. If we had insisted on a minimum of six daily measurements per month, we would have had only 68 monitors available in the 84-month sample. Using a 72-month sample increased the available monitors by a factor of two or more.

Because of the relative scarcity of monitors with complete data, we examined whether it would make sense to use daily data from nearby monitors to predict a missing reading. For this procedure to be useful, a monitor must have neighbors, and the readings from the neighbors must be strongly correlated. For each monitor for which it was possible, we regressed its daily readings for each calendar year against the average daily (same-day) readings of other monitors in the county. Not surprisingly, the spatial correlation of these monitors was quite high. The average r^2 for the 2870 available monitor-years was 0.85, and the interquartile range was 0.81 to 0.95. To be sure, some r^2's were very low, including 13 monitor-months with r^2's of less than 0.2. In each of these cases, the year with the low reading was an anomaly, and the r^2's for the other years at the same monitors were much higher. This suggested that for most monitors it would be useful to supplement the data with readings from other monitors in order to avoid the need to discard a potentially large amount of valuable information. Specifically, then, for each monitor and each missing month, we added daily data from other monitors in the same county on that day, taking the mean daily observation if there was more than one other monitor with readings on that day. Table 8 replicates Table 7 with the monitor data enhanced by the other data as described above. Comparing the two, some improvement in sample size is evident, and coverage is significant. However, this was no panacea, as the number of available monitors grew by only 10% to 15%.

For our estimation we only accepted monitors with readings on at least two different days in each of the 84 months in the 1999–2005 study period. This requirement imposed a serious restriction on the number of usable monitors. As shown in Table 8, only 193 monitors in the eastern United States met the criterion, giving us a sample size of 193 × 84 = 16,212 observations. If we restricted the time interval to the years 2000–2005, we found 413 available monitors, yielding a sample size of 29,736 observations.

Table 9 shows descriptive statistics for the 193 monitors in our main dataset.

Table 7. Tradeoff Between Monitors Available and Minimum Monitoring Frequency: Initial Datasets

Minimum Monthly Monitoring Frequency	Monitors Available	
	1999–2005 Dataset	2000–2005 Dataset
1	204	412
2	168	381
3	141	344
4	113	300
5	90	253
6	68	202

Table 8. Tradeoff Between Monitors Available and Minimum Monitoring Frequency: Enhanced Datasets

Minimum Monthly Monitoring Frequency	Monitors Available	
	1999–2005 Dataset	2000–2005 Dataset
1	224	442
2	193	413
3	165	391
4	138	349
5	112	305
6	85	254

Table 9. Characteristics of 1999–2005 Monthly $PM_{2.5}$ Dataset[a]

Characteristic	Monthly Averages		
	Mean	Maximum	Minimum
Average monthly temperature (°K)	287.8	306.1	259.0
Average monthly $PM_{2.5}$ ($\mu g/m^3$)	14.5	39.1	3.4
Average number of daily observations for mean computations per site	14.2	30.4	4.6

[a] Number of monitors = 193; number of months per site = 84; total monthly observations = 16,212.

STATISTICAL MODEL

In this section, we describe the statistical approach we used to link emissions and ambient $PM_{2.5}$. We ran a number of alternative linear regression models that included extensive fixed-effects variables to capture mobile and area source emissions and other site-specific factors resulting in variation in $PM_{2.5}$ that could not be observed in the necessary temporal and spatial detail. These models differed in the functional form of the emission variables and in the structure of the fixed-effects variables. The initial model included the source emissions and monitor-site temperature plus fixed effects for the monitor site, year, and month, as shown here:

$$A_{it} = \sum_{j=1}^{8} \sum_{d=1}^{k} \beta_{jd} X_{it}^{jd} + \tau(T_{it} - \bar{T}_{im}) + \sum_{i'=1}^{N} \sum_{y'=1}^{Y-1} \alpha_{iy} R_{it}^{i'} Y_{y}^{y'} + \sum_{m'=1}^{11} \phi_m M_m^{m'} + \varepsilon_{it}, \quad (1)$$

where i and t indicate the monitor and sequential month, respectively, with $i = 1, \ldots, N$ and $t = 1, \ldots, T$; and y and m indicate the calendar month. The sequential counter t and the year and calendar-month indicators y and m are related by the expression $t = 12(y - 1) + m$; the first double summation on the right-hand side represents the emissions in zone j and direction d from monitor i. (The specification of the emission variables will be discussed in some detail below, in the following section.) $A_{it} = A_{iym}$ represents the monthly average of daily $PM_{2.5}$ readings; $T_{iym} - \bar{T}_{im}$ is the deviation of the mean monthly temperature from the mean averaged over the 1999–2005 time period; α_{iy} and ϕ_m represent fixed effects for the interacted monitor and year variables (described below) and month, respectively; and ε_{iym} represents the disturbance terms.

The monitoring data we used covered 193 monitors over the 7-year period from the beginning of 1999 to the end of 2005, as described above, for a total of 16,212 observations.

Emissions

The simplest way to organize our emissions data would have been to treat each source as an independent variable in our model. However, the resulting number of variables (1314) would have been unnecessarily large, inasmuch as sources in close proximity to one another likely have quite similar impacts on distant receptors. Accordingly, we opted for what we believe to be a plausible emission aggregation strategy. As shown in Figure 1, we divided the area around each monitor into a set of zones defined by the intersection of, in this example, three concentric rings (representing three distances) centered on the monitor and eight wedges bounded by the principal points of the compass (north, northeast, east, etc.). X_{it}^{jd} is the sum of SO_2 and NO_x emissions from all utility sources located in wedge j and ring d for monitor i during month t. As shown in Figure 2, emissions from a source could be associated with many different monitors, depending on the direction and distance to each. If the outermost zone in figures like these is defined as extending to an infinite distance, then emissions from every source would be associated with every monitor. This would not necessarily be true if the outer zone were defined as extending only to a finite distance.

The emissions values in equation 1 are for SO_2 and NO_x emissions from electric power generators, which are subject to monitoring by continuous emission monitors. Both

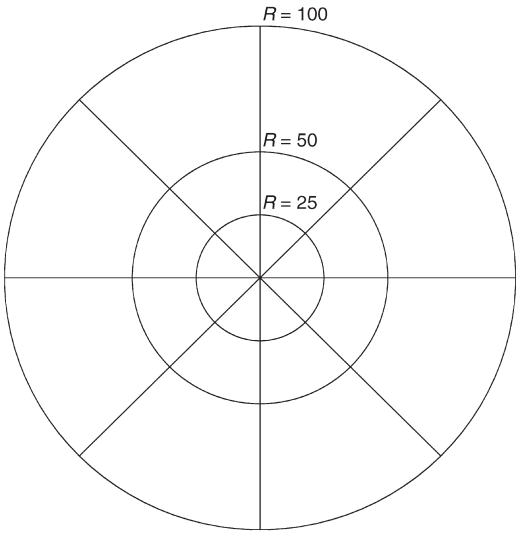


Figure 1. Example of emission zones around a monitor. *R* indicates radius in miles.

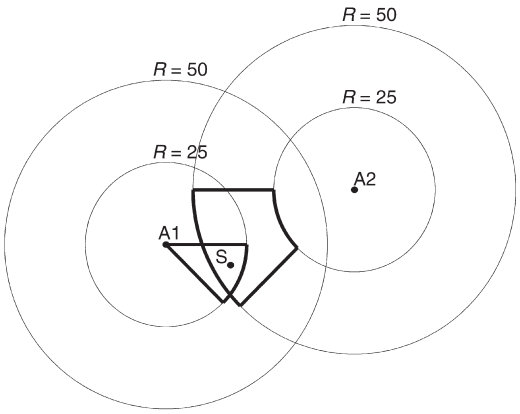


Figure 2. Example of overlapping relationships between a source (S) and various monitors (A1 and A2). *R* indicates radius in miles.

of these pollutants are precursors of $PM_{2.5}$. However, a serious difficulty arises when both pollutants are entered in the equation simultaneously, because at each source the SO_2 and NO_x emissions are highly correlated with each other. Figure 3 illustrates the difficulty. For an area divided into, in this case, 64 zones (defined by the intersection of eight concentric rings [representing eight distances] and eight wedges [representing eight directions]), the average within-zone correlation of NO_x and SO_2 emissions is 0.88, and only three observations have correlations below 0.8.

Fortunately, within-zone emissions of SO_2 or of NO_x taken individually do not cause this difficulty in the models, or at least not as seriously. As shown in Figure 4, the multicollinearity (i.e., degree of correlation) of SO_2

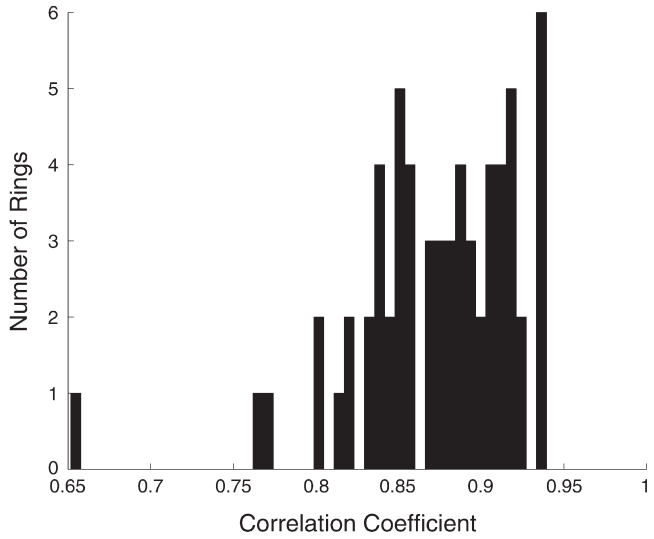


Figure 3. Multicollinearity of NO_x and SO_2 emissions.

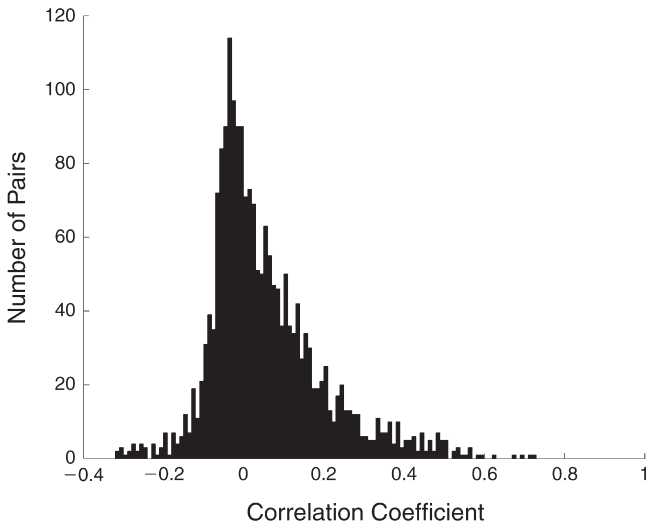


Figure 4. Multicollinearity of SO_2 emissions across zones.

emissions across zones is much lower; only a few of the 2016 zone pairs in this configuration have correlation coefficients that exceed 0.6 in absolute value. However, the variables involved bear watching, because it only takes one such pair to cause the matrix of independent variables to be ill conditioned. Fortunately, the zones with highly correlated SO_2 and NO_x emissions tended to be those with few sources and many nonzero entries, with the result that the corresponding variables could be dropped from the regression equation without creating an omitted-variable problem.

For each of our zone specifications, we not only entered the emissions in tons, but also created an interaction variable using the emission variable and the average temperature at the monitor — that is, we calculated a second specification for the model, as follows:

$$A_{it} = \sum_{j=1}^{8}\sum_{d=1}^{k}\beta_{jd}X_{it}^{jd}T_{it} + \tau(T_{it} - \bar{T}_{im}) + \sum_{i'=1}^{N}\sum_{y'=1}^{Y-1}\alpha_{iy}R_{it}^{i'}Y_{y}^{y'} + \sum_{m'=1}^{11}\phi_{m}M_{m}^{m'} + \varepsilon_{it}. \quad (2)$$

Note that when we interacted temperature with emissions we also entered temperature alone. Within this basic structure of emission types and zone definitions, we then examined the following:

- A range of specifications that included both a full set of SO_2 and NO_x variables in the equation regardless of the multicollinearity problem;
- A range of specifications that included only SO_2 emissions, keeping in mind that the SO_2 emissions were representing both SO_2 and NO_x emissions; and

- A range of specifications that included a weighted sum of SO_2 and NO_x emissions, varying the weights and observing the effect on the mean squared error (MSE) of the model and the plausibility of the individual coefficients.

We found that all of these specifications were inferior to or produced results that were very similar to those of simply using SO_2. We therefore focused on SO_2 in the reported results, understanding that the SO_2 variable also included the effects of NO_x. Note that in Phase 2 of the Title IV program an average of approximately 0.8 tons of NO_x was reduced for every ton of SO_2. (The SO_2:NO_x ratio is relevant for the comparisons with CAIR made later in the report. In CAIR, this ratio was considerably lower [less than half] than that in Title IV Phase 2.)

Our specifications for defining the emission zones included the following:

- Varying the number of rings;
- Varying the orientation and number of wedges (we experimented with four quadrants or even two halves in addition to our usual practice of dividing the area into eighths).

The main effects of changing the zones were found in changing the number of rings. Changing the orientation and number of the wedges had only small effects.

Temperature

Our temperature variable T was the ground-level Kelvin temperature at or near each monitoring site, as recorded in the EPA's Air Quality System. To reduce possible correlations with other variables, we defined T as the difference between the reported monthly temperature and the average monthly temperature over the 7-year period at the same monitor for the same month.

Temperature was the only meteorologic variable used in the analysis. Initially we had planned to use wind speed and direction data to condition the source emission coefficients. But after estimating, with little or no success, a variety of specifications that made use of wind data, we concluded that the atmosphere was much too complicated and had too much spatial variability for us to capture the interactions of wind data with the emissions–air quality relationship, especially for monitors located hundreds of miles away. Nonetheless, it should be noted that our dartboard-like zone structure for emissions incorporated, at least in a rudimentary way, one of the chief meteorologic variables likely to affect emissions transport, namely the prevailing west-to-east wind patterns. We expected to see coefficients for the emissions zones to the west of each monitor to be larger than the coefficients for zones to the east.

Fixed Effects

There are many local influences on site-specific $PM_{2.5}$ concentrations, including emissions from motor vehicles and area sources; geospatial characteristics, such as terrain features and the presence of nearby bodies of water; local weather; and many more. For most of these we had limited data or none at all. If these effects are ignored, the estimated coefficients will likely suffer from omitted-variable bias, a quite common problem with panel data (Hsiao 1986).

To avoid this bias and make up for the dearth of data, researchers often incorporate random or fixed effects in their models. Random effects — random disturbance terms that are entered only for some observations — are more efficient, but as error terms they are assumed to be uncorrelated with any of the other independent variables. Any correlation leads to potential bias in the estimated coefficients of those variables. Fixed effects are additional independent variables, equivalent to a model estimated on mean deviations. They are logical variables designed to account for unobserved heterogeneity. They take on a value of 1 or 0 for each observation depending on whether the observation is true or not for that variable. Thus, for example, our site fixed effect was defined by:

$$R_{iym}^{i'} = \begin{cases} 1 & \text{if } i' = i \\ 0 & \text{otherwise}. \end{cases}$$

The fixed effects for year and month were defined similarly. They sacrificed efficiency, but unlike random effects they do not introduce bias unless the mean effect of the variable changes over the set of observations for which it is true. As shown by the second double sum in equation 1, all of our specifications included interacted fixed-effects variables for year and site. This allowed us to take into account not only local effects for which we did not have reliable monthly data (e.g., the effects of terrain and local weather and the area-average local effects of mobile or area emission sources), but also local and national trends.

Fixed effects add significantly to the number of parameters that must be estimated, especially when terms are interacted. In equation 1, for example, we had 1351 interaction terms for site and year.

Seasonal Effects

The inclusion of fixed effects for month in equations 1 and 2 were intended to account for seasonal effects on ambient $PM_{2.5}$ that could not be captured by the year and site dummy variables described above. However, these monthly effects are strongly related to temperature and to the effects of emissions on $PM_{2.5}$ concentrations — exactly what we were trying to observe. We therefore estimated our models (equations 1 and 2) both with and without the

monthly variables and observed the effects on the coefficients. In specifications without monthly variables, the temperature was entered as recorded, not as a deviation from the monthly mean, as in equation 3:

$$A_{it} = \sum_{j=1}^{8} \sum_{d=1}^{k} \beta_{jd} X_{it}^{jd} T_{it} + \tau T_{it} + \sum_{i'=1}^{N} \sum_{y'=1}^{Y-1} \alpha_{iy} R_{it}^{i'} Y_{y}^{y'} + \varepsilon_{it}. \quad (3)$$

Disturbance Term

The standard linear regression model, known as ordinary least squares (OLS), makes two strong assumptions about the nature of the random disturbance term ε_{iym}.

- Assumption 1 is that the term is homoskedastic — i.e., that all errors are normally distributed, with mean zero and a common variance σ^2. It is noted as $\varepsilon_{iym} \sim N(0,\sigma^2)$.
- Assumption 2 is that errors are uncorrelated over space and time — i.e., that $\text{cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0$ if $i \neq j$ and $\text{cov}(\varepsilon_{it}, \varepsilon_{is}) = 0$ if $t \neq s$.

When these assumptions are violated, OLS estimation produces estimated coefficients that are not efficient. Furthermore, if the errors are correlated in space or time, then the estimated covariance matrix is biased. Examination of the residuals for a typical OLS specification showed them to be heteroskedastic and correlated over both space and time, as shown in Figures 5 through 7. These results were typical for all OLS specifications of the model. Violation of the OLS assumptions of constant variance and no contemporaneous spatial correlation is quite serious; temporal correlation is less so.

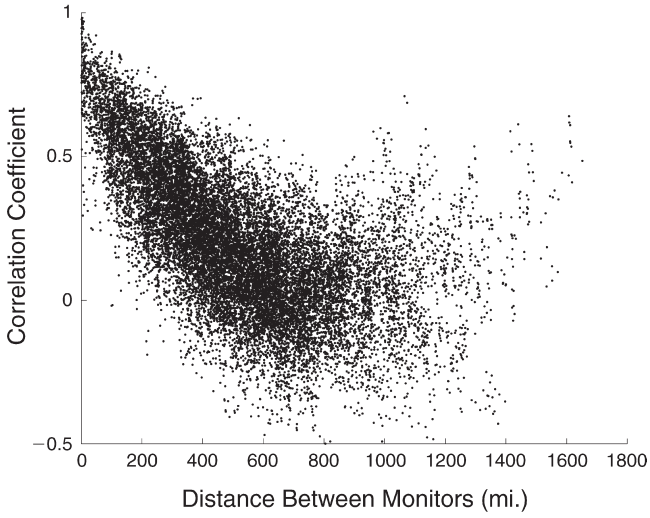


Figure 6. OLS error correlation versus distance.

Dealing with Departures from the Ordinary Least Squares Assumptions

For ease of reading, we used vector notation for the dependent variable $\mathbf{A}_i$ and generic matrices $\mathbf{Z}_i$ to represent the variables in the right-hand side of equations 1, 2, and 3 and rewrote equation 1 more simply in block form as:

$$\mathbf{A} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_N \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix}, \quad (4)$$

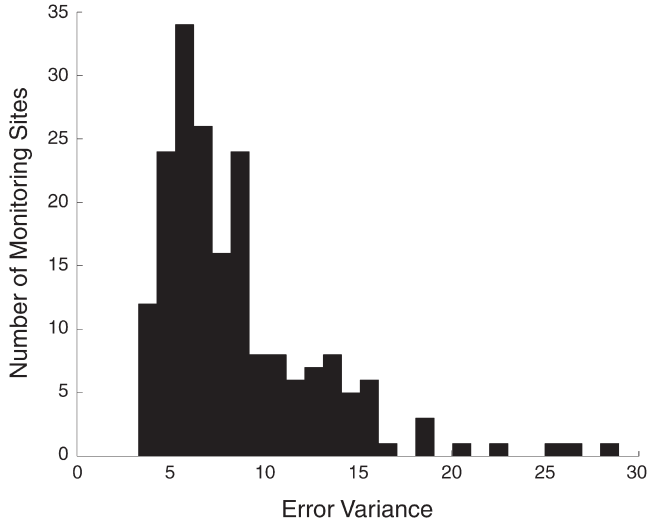


Figure 5. Distribution of error variances.

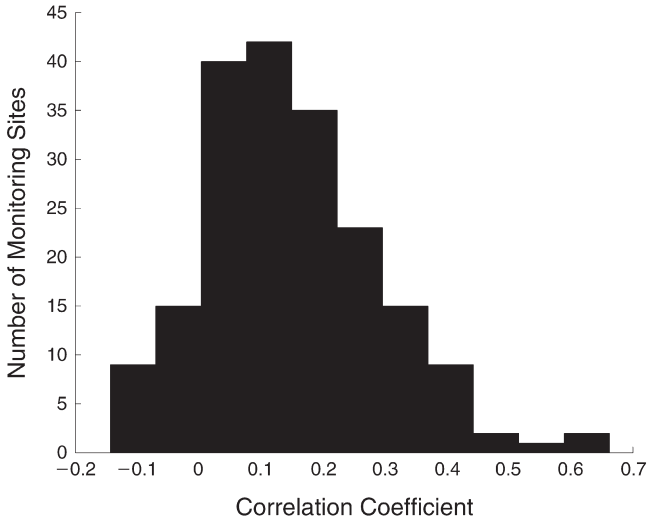


Figure 7. Autocorrelation of OLS errors.

where each block vector or matrix corresponds to a monitor site with T observations. In matrix terms, the OLS assumptions implied that the covariance matrix can be written as $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $NT \times NT$ identity matrix.

We needed to relax the OLS assumptions in order to address the issues of contemporaneous spatial correlation, autocorrelation, and heteroskedasticity. In their place we adopted the following assumptions:

Contemporaneous Spatial Correlation $cov(\varepsilon_{it}, \varepsilon_{ju}) = \sigma_{ij}$.

Serial Correlation We assumed an AR(1) process for the errors, so that $cov(\varepsilon_{it}, \varepsilon_{i,t-j}) = \rho_i^j$.

Heteroskedasticity The OLS assumption of constant error variance is an especially strong one, because there are numerous ways that nonconstant variances can arise. Here we took into account two of the most important:

1. Group heteroskedasticity. We assumed $\varepsilon_{it} \sim N(0,\sigma_i^2)$, so that the estimation errors at each site would have the same variance, but the variance of errors at difference sites could be different.
2. Sampling frequency. Average monthly $PM_{2.5}$ concentrations at each monitor were estimated from readings that were supposed to be taken at regular intervals — daily, every 3 days, or every 6 days — but often were not. This meant that the number of daily readings in a month could vary from 1 to 31; for that reason the precision of monthly estimates differed greatly for different months even at the same site.

We deal, below, with sampling frequency first and then everything else in a single step.

Sampling Frequency As discussed earlier, most of the monitors in the sampling network had periods when very few daily measurements were being made. To obtain a balanced panel, we were forced to use such monitors anyway; if we had not, we would have had very few monitors with which to do our analysis. Rather than discard a monitor, we realized, it was better to correct for the months with only small numbers of daily observations in them by discounting their importance when estimating a model.

This correction was based on the well-known formula for the variance of a sample mean. If $A_{iym} = \Sigma d_{iymj}/N_{iym}$ in equations 1, 2, and 3 is the mean of N_{iym} daily observations, which we assumed to be independent and identically distributed, then $\text{var}(A_{iym}) = \text{var}(d_{ij})/N_{iym}$ is the sampling variance of the A_{iym}. We could correct for the nonconstant sampling variance by weighting the observations by either the monthly variance directly or by the number of observations for the month. It turned out that weighting by the number of monthly observations was far superior, because it guaranteed that months at individual monitors with small numbers of observations were given small weights. Using the monthly variance gave too much weight to months with small numbers of observations that happened to be very close in magnitude.

In the first step of the estimating procedure, therefore, we transformed equation 4 to:

$$\sqrt{\mathbf{N}}.*\mathbf{A} = \sqrt{\mathbf{N}}.*\mathbf{Z} + \sqrt{\mathbf{N}}.*\boldsymbol{\varepsilon}, \tag{5}$$

where $\sqrt{\mathbf{N}}$ is a column vector containing the number of daily observations represented in each monthly estimate, and .* is the operation of pairwise multiplication.

Correcting for Contemporaneous Spatial Correlation, Serial Correlation, and Group Heteroskedasticity If $\mathbf{Z}$ represents the complete matrix of independent variables in equation 1 and $\hat{\gamma}$ represents the vector of coefficients estimated by OLS, then the errors $e_{it} = A_{it} - \|\mathbf{Z}\hat{\gamma}\|_{it}$ for $t = 1, \ldots, T$. The error structure of equation 4 is now:

$$\mathbf{V} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \cdots & \boldsymbol{\Omega}_{1N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}_{N1} & \cdots & \boldsymbol{\Omega}_{NN} \end{bmatrix}, \tag{5}$$

where

$$\boldsymbol{\Omega}_{ij} = \frac{\sigma_{ij}}{1 - \rho_i\rho_j} \begin{bmatrix} 1 & \rho_j & \rho_j^2 & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & \rho_j & & \rho_j^{T-2} \\ \rho_i^2 & \rho_i & 1 & & \vdots \\ \vdots & & \ddots & & \\ \rho_i^{T-1} & \cdots & & \rho_i & 1 \end{bmatrix}. \tag{6}$$

OLS could not be used to estimate the models (i.e., equations 1, 2, and 3) with the error structure in equations 5 and 6. Instead, we used feasible generalized least squares (FGLS) and maximum likelihood (ML) (Greene 1993). These estimation methods are asymptotically equivalent but can give very different results in small samples. With our relatively large dataset, we found that the results for the two methods were quite similar.

The generalized least squares estimator for the model was:

$$\tilde{\boldsymbol{\beta}} = (\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{A}. \tag{7}$$

To estimate equation 7, we needed estimates for the parameters in each $\boldsymbol{\Omega}_{ij}$ in equation 6. The basic strategy was to estimate equation 1 using OLS and to use the OLS residuals to estimate the site-level error variances σ_i^2 and

the autocorrelation coefficients ρ_i. Denoting the OLS residuals by e_{it}, our estimates of these parameters were:

$$s_{ii} = s_i^2 = \frac{1}{T}\sum_t e_{it}^2 \quad \text{and} \quad r_i = \frac{\sum e_{it}e_{i(t-1)}}{(T-1)s_i^2}, \tag{8}$$

respectively.

Contemporaneous Spatial Correlation

Unfortunately, we could not use a similar method to obtain estimates of the covariance terms in equation 6, because too many covariances were required. Although we could compute the sample covariance for any two sites by $s_{ij} = \Sigma e_{it}e_{jt}/T$ (as we did to generate Figure 6), the matrix $\|s_{ij}\|$ would not be of full rank. With 193 monitors and 84 months of data, we would have had 18,528 covariance terms to estimate and only 16,212 observations. We used the fact that the observed correlations were strongly related to distance, as illustrated in Figure 6, to estimate a simple model of covariance. That is, we specified that the error covariance between sites i and j depended on the error variances at each site and an exponential function of the distance $D(i,j)$ between them, and we estimated this covariance σ_{ij} by:

$$s_{ij} = s_i s_j \exp(-\alpha^2 D(i,j)). \tag{9}$$

We then estimated the decay coefficient α by finding the best least squares fit to the site-level empirical correlations shown in Figure 6.

The expressions in equations 8 and 9 gave us everything we needed to produce a FGLS estimator of the $\boldsymbol{\beta}$. The same expressions also provided the starting values for estimates of the σ's and ρ's in the maximum likelihood estimation (MLE).

Using Estimated Coefficients to Generate Counterfactuals

To generate a set of n random draws from a common normal distribution $N(\mu,\sigma^2)$, we drew a set of n standard normal variables u_i, where $u_i \sim N(0,1)$ and the desired set of random draws is $x_i = \mu + \sigma u_i$. For a multivariate normal random variable coefficient vector $\tilde{\boldsymbol{\beta}}$, as in equation 6 above, the procedure is analogous.

The parameter vector $\tilde{\boldsymbol{\beta}}$ has the distribution $N(\boldsymbol{\beta}, var(\tilde{\boldsymbol{\beta}}))$, where $var(\tilde{\boldsymbol{\beta}})$ is the covariance matrix of $\tilde{\boldsymbol{\beta}}$. Because $var(\tilde{\boldsymbol{\beta}})$ is definitely positive, it can be written as $\mathbf{C\Lambda C}^T$, where $\boldsymbol{C}$ is a matrix whose columns are the characteristic vectors of $var(\tilde{\boldsymbol{\beta}})$ and Λ is a matrix that has the (all positive) characteristic roots along the diagonal and zeros elsewhere. Now if we draw a standard normal vector $\mathbf{u}$, $\mathbf{u}^T = (u_1, u_2, \ldots, u_k)$, with the same dimensionality as $\tilde{\boldsymbol{\beta}}$, then the vector $\mathbf{x} = \boldsymbol{\beta} + (\mathbf{C}\sqrt{\boldsymbol{\Lambda}})\mathbf{u}$ has the required distribution $N(\boldsymbol{\beta}, var(\tilde{\boldsymbol{\beta}}))$.

MODEL RESULTS

This section presents our regression results in two parts. The first subsection is primarily descriptive, focusing on the fixed-effects variables described in equations 1 and 2 earlier to examine the importance of local and temporal factors as determinants of $PM_{2.5}$ concentrations. In this initial model specification, we did not include power plant emissions as an explanatory variable. The second subsection introduces SO_2 and NO_x emissions from power plants as explanatory variables and considers a range of model specifications. Additionally, as a sensitivity analysis, we subdivide the sample to test for regional differences. Given the large number of explanatory variables, we focus attention in this section on relatively simple tabular and graphical presentations.

Analysis of Variance of $PM_{2.5}$ Concentrations

Here we will conduct a relatively simple descriptive analysis before entering emissions into the equation — essentially an analysis of variance. For ease of presentation we divided this simple analysis into two parts. In the first part we considered three sets of variables one after another — site, year, and month — and calculated the sample variance around the site, year, and month means, respectively. For example, if A_{iym} is the monthly observation at site i, year y, and month m, the site variance is defined by:

$$var_I(A) = \sum_{y,m} (A_{iym} - \bar{A}_{.ym})^2.$$

This site variance is equivalent to the MSE of a statistical model, estimated by OLS, with a full set of dummy variables for each monitor site. The year and month variances were defined analogously. As shown in Table 10, the variance by year (17.60) was only slightly smaller than the overall variance of the sample (18.33), but the average site-specific (12.95) and month-specific (12.73) variances were smaller by almost one-third.

We next considered the two-way combinations, that is, the month–year, site–year, and site–month interactions. The site–year variance, for example, is defined by:

$$var_{IY}(A) = \sum_m (A_{iym} - A_{..m})^2.$$

Again, this variance is equivalent to the MSE of an OLS regression of ambient $PM_{2.5}$ against a full set of 1351 dummy variables representing combinations of site and year. Not surprisingly, the two-way variances for site and year were smaller than either of the respective one-way variances in Table 10. For the interactions involving year, the reduction in variance over the one-way variances was not large. But the site–month interactions reduced the

Table 10. Mean Squared Errors in Models with Only Fixed Effects[a]

	1999–2005	
	Parameters	MSE
Overall	1	18.33
One-way models		
Site dummies	193	12.95
Monthly dummies	12	12.73
Annual dummies	7	17.60
Two-way models		
Site–year interactions	1351	11.64
Site–month interactions	2316	7.23
Month–year interactions	84	12.73

[a] Number of monitors = 193; total monthly observations = 16,212.

Table 11. Average of Mean Squared Error (Using FGLS) at Each Monitor

	Emission Zone Radii		
	0–300	0–400	0–500
Without Monthly Dummies			
TEMP ONLY	11.67	11.67	11.67
TEMP + EMIS	9.81	9.71	9.76
TEMP × EMIS	9.54	9.46	9.53
TEMP × EMIS (MLE)	9.68	9.64	9.74
TEMP × EMIS (North)	10.03	10.03	10.21
TEMP × EMIS (South)	8.12	7.86	7.76
TEMP × EMIS (NO_x)	11.28	11.11	11.08
With Monthly Dummies			
TEMP ONLY	10.27	10.27	10.27
TEMP + EMIS	9.45	9.55	10.00
TEMP × EMIS	9.30	9.37	9.82
TEMP × EMIS (MLE)	9.47	9.53	10.07
TEMP × EMIS (North)	9.06	9.15	9.78
TEMP × EMIS (South)	8.27	8.00	7.78
TEMP × EMIS (NO_x)	10.41	10.47	10.71

average of the MSEs, to 7.23 μg/m^3, thereby explaining almost 61% of the total variance in the $PM_{2.5}$ concentrations. As in the case of the simple introduction of the monthly dummy variables described above, however, we believe that the monthly variables were picking up the influence of weather and, most important, emissions — factors that we explicitly consider in the next section.

Models Linking Power Plant Emissions to $PM_{2.5}$ Concentrations

This section reports the results of the statistical models described above. Once again, the emission sources associated with a monitor were grouped like the segments of a dartboard, as shown in Figure 1. The average of the MSEs reported here reflects three different ring definitions for each monitor. The first set of results, shown in the first column of Table 11, defined the rings to include the following distances (in miles) of the monitors from the sources: 0–12.5, 12.5–25, 25–50, 50–100, 100–150, 150–200, and 200–300. The second and third sets of results add the sources in the ring 300–400 and 400–500 miles, respectively, from the monitors. The average emissions in each of the rings, shown in Table 12, varied from one monitor to another.

In all models, the disturbance terms were allowed to be autocorrelated and heteroskedastic, with contemporaneous covariance based on distances between monitor sites. Given these very substantial departures from the OLS assumptions, we needed to use the alternative estimating methods of FGLS and ML. We assessed model performance by the criterion of average MSE from the individual site means, calculated using the original dependent variables rather than the transformed variables required by the estimation procedures. By this criterion, we found that FGLS and ML gave roughly comparable results. Both were asymptotically efficient; in finite samples they could yield different results, and neither dominated the other. We have primarily reported FGLS results here, because the estimation procedure was less time-consuming and the MSEs were slightly smaller.

Table 12. Average Monthly Emissions in Rings Around Monitors

Radius (Miles)	Average Monthly Emissions (Kilotons)
0–12.5	176
12.5–25	273
25–50	831
50–100	2,639
100–150	3,915
150–200	4,285
200–300	12,717
300–400	15,982
400–500	15,228
>500	51,770

The MSEs of a range of alternative models for three emission-zone specifications and 14 independent-variable specifications are shown in Table 11 (except one using FGLS). These specifications all used the interacted site–year fixed effects described above plus various formulations of temperature and emissions variables. The results are shown with and without the monthly dummy variables. The monthly variables are highly significant and were obviously correlated with the emission and temperature variables we cared about, meaning that to leave them out would bias the coefficients on the other variables. However, we believe that comparison of individual emission and temperature coefficients for models with months and models without months was useful for the interpretation of the results. The seven alternative variable specifications were the following:

- **TEMP ONLY** The (average monthly Kelvin) temperature
- **TEMP + EMIS** Temperature and emission variables, entered additively, for the indicated emission zones
- **TEMP × EMIS** Emission variables interacted with temperature (which was also entered alone)
- **TEMP × EMIS (MLE)** Estimation by maximum likelihood
- **TEMP × EMIS (North)** Estimation limited to 116 northern monitors (shown in Figure 8)
- **TEMP × EMIS (South)** Estimation limited to 77 southern monitors (also shown in Figure 8)
- **TEMP × EMIS (NO_x)** NO_x emissions replacing SO_2 emissions

Examination of Table 11 suggests the following observations:

- Addition of the emissions in the outermost zone (400–500 miles) seldom improved the fit of the model and in some cases actually increased the MSEs. Only for the southern monitors did the addition of the outermost zones' emissions seem to improve the MSEs.
- Inclusion of monthly dummy variables significantly reduced the MSEs in most cases.
- For temperature alone, the average of the MSEs was approximately the same as — actually slightly higher than — the site–year variance shown in Table 10 without the temperature variable. (The MSEs were higher because the calculations for Table 10 were made under the more restrictive OLS assumptions and no doubt underestimated the true variance.)
- Adding the emission variables significantly reduced the MSEs compared with temperature alone.

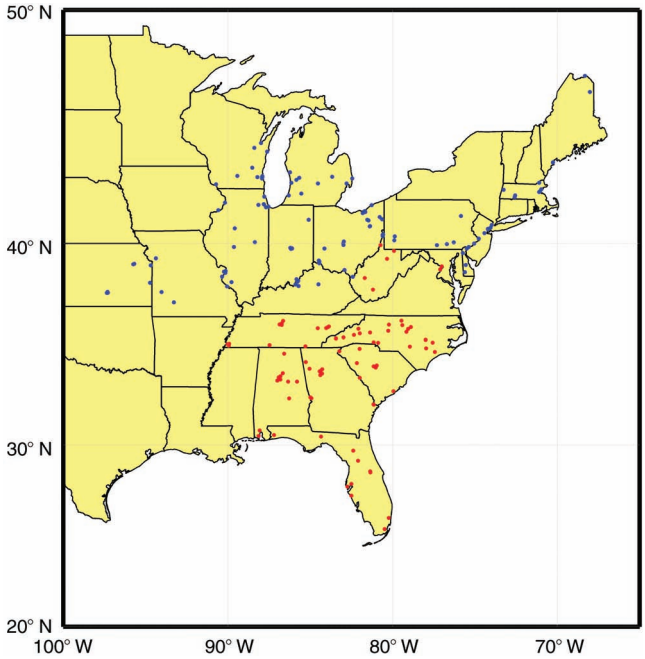


Figure 8. Locations of monitors in northern (blue) and southern (red) regions.

- Interacting temperature and emissions outperformed emissions entered linearly.
- MSEs were higher for the northern monitors than for the southern ones, perhaps suggesting that what we called the southern region was somewhat more homogeneous and thus more amenable to being represented by a single set of emission coefficients.
- There seems to be little basis for choosing between the models with and without the 300-to-400-mile ring. Neither dominated for all specifications.
- Substituting NO_x for SO_2 significantly reduced model performance. (Other, mixed specifications [not shown], including weighted sums of NO_x and SO_2, also performed worse than SO_2 alone.)

Characteristics of Regression Coefficients

In this section we describe the characteristics of the individual regression coefficients.

The first thing to note is that the results when temperature was interacted with the emission variables (TEMP × EMIS) were actually quite similar to the results when the emission variables were entered without interaction (EMIS), once an adjustment was made for the difference in units. Figure 9 plots the emission coefficients for TEMP × EMIS (with monthly dummy variables) multiplied by the

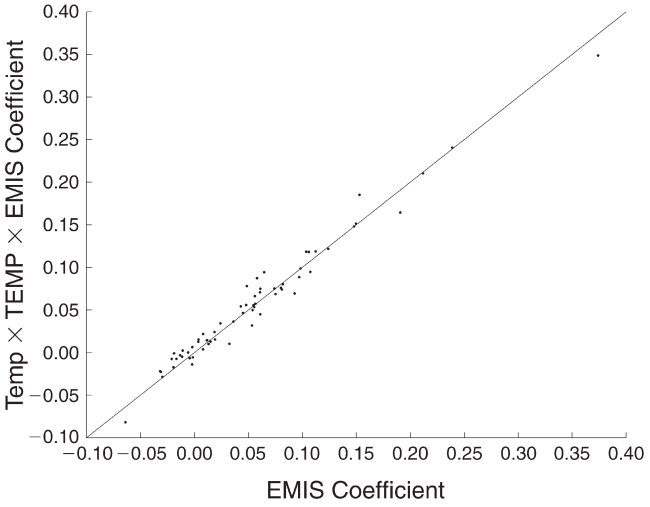


Figure 9. Coefficients for TEMP × EMIS (with monthly dummies) multiplied by the average temperature for the sample plotted against coefficients for EMIS for the same model specifications.

average temperature for the sample against the EMIS coefficients. As shown, the coefficients line up very close to the 45-degree line. We preferred the models with the temperature–emission interaction, because they were more sensitive to actual temperature rather than to the mean temperature of the sample, with the result that they tended to have slightly lower MSEs.

The important characteristics of the emission coefficients can be summarized as follows:

- Although there were no restrictions on the sign of the emission coefficients, a large majority were positive. This was especially true in specifications without monthly variables. In specifications with monthly variables, we found more negative signs, but — except in the outermost zone — the negative coefficients were insignificant.
- Beyond 25 miles, the value of the emissions–concentrations gradients generally declined as a function of distance from the source. This relationship held over the full distance examined, out to about 400 miles. MSEs also declined with distance from the source, probably reflecting the large number of zero values for emissions in the closer-in, and thus smaller, emission zones.
- Within 25 miles, there was generally a peak in the value of the emissions-concentrations gradients. The exception was for eastern sources in the northern region, where there was a more continuous decline in the emissions-concentration gradients out to about 75 miles.
- Consistent with the prevailing westerly winds, the values of the western sources tended to be higher than those of the eastern sources. Nonetheless, the differences were often not statistically significant, and many of the coefficients for emission zones east of monitors were statistically significant and of sizable magnitude.

Figures 10 to 15 show the source–receptor coefficients (SRCs), based on the distance between sources and monitors. Most of the results were for models that excluded dummy variables for month, although the coefficient plots when months were included were very similar. Figure 10 shows the results of the third model considered in Table 11, that is, with the inclusion of the site–year fixed effects and the temperature and emission variables, using the full sample and FGLS, focusing on the average effect of the distance of the source (SO_2) emissions on $PM_{2.5}$ concentrations. (The mean distance from a monitor of a ring bounded by radii R_1 and R_2 is $\sqrt{(R_1^2 + R_2^2)/2}$.) The gray line shows the average effect for all sources located in the four zones east of the monitor. The results for individual zones are represented by gray diamonds. The black line shows the average effect for all sources located in the four zones west of the monitor. The black squares show the results for the individual zones. At each distance the values of the individual coefficients are shown together with their error bars, defined at two SDs. For ease of viewing, we have separated the two sets of error bars; in fact they should be superimposed.

As shown, the largest impact (per unit of emissions) occurred at a distance of about 25 miles from the source.

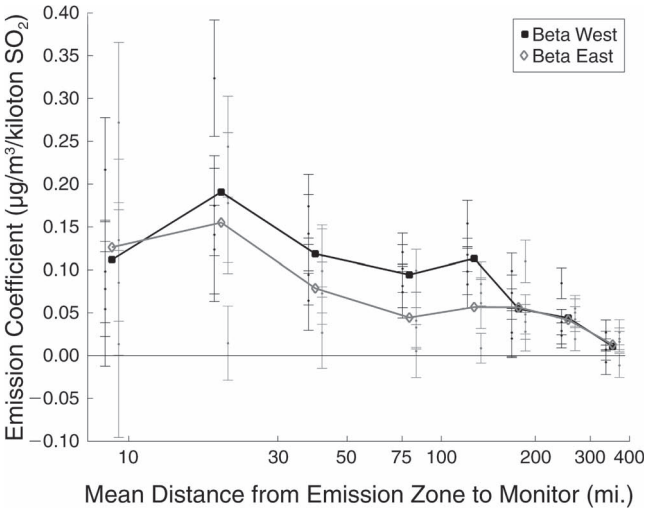


Figure 10. Estimated effects of emissions on ambient $PM_{2.5}$ by distance from source (logarithmic scale), for all 193 monitors.

For emissions east of the monitor, we found that as the distance from the emission source increased there was less of an impact from emissions reductions of $PM_{2.5}$ concentrations. Beyond about 25 miles from the source, the effect of SO_2 emissions on $PM_{2.5}$ concentrations declined almost monotonically out to about 80 miles. At that point we saw a slight increase in the SRCs (well within the error bounds), and then another monotonic decline. Overall, there was clearly an inverse relationship between distance and the effect on concentrations caused by emissions reduction. Not surprisingly, western sources generally had larger impacts on $PM_{2.5}$ concentrations than eastern sources, although the differences were not statistically significant. Note also that the error distributions were wider for the eastern sources at the close-in distances; this pattern did not persist over longer distances.

Figure 11 shows the results for the same model as that used for Figure 10, except that the sample was restricted to the 116 northern monitors in our dataset. As can be seen, the pattern of results was roughly comparable to that of the full sample, albeit with much larger standard errors. At the same time, there were some important differences between the two figures. Most important, whereas for western sources we saw a peak in the effect of emission sources on $PM_{2.5}$ concentrations at about 25 miles (as in the full sample), there was no such peak for eastern sources, for which the effects declined continuously out to about 75 miles. At that point we did see an increase for eastern sources, although the standard errors were quite large.

Figure 12 shows the results for the 77 southern monitors in our dataset. As with the results for the northern subsample, the standard errors were somewhat larger than those of the full sample. In this case, both the eastern and southern sources had a peak at about 20–25 miles, followed by reasonably monotonic declines thereafter.

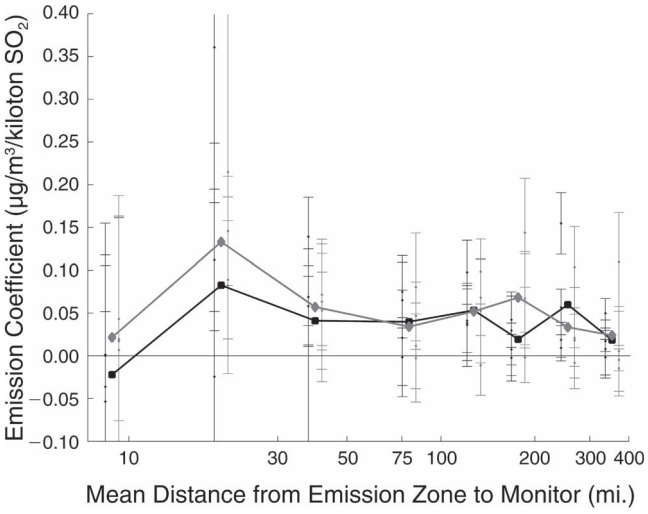


Figure 12. Estimated effects of emissions on ambient $PM_{2.5}$ by distance from source (logarithmic scale), for 77 southern monitors.

We now turn to a consideration of the alternative models, which included fixed effects for months as additional independent variables. As shown in Figures 13 to 15, which show results for all 193 monitors, the 116 northern monitors, and the 77 southern monitors, respectively, the pattern of the SRCs over the area within 400 miles of a

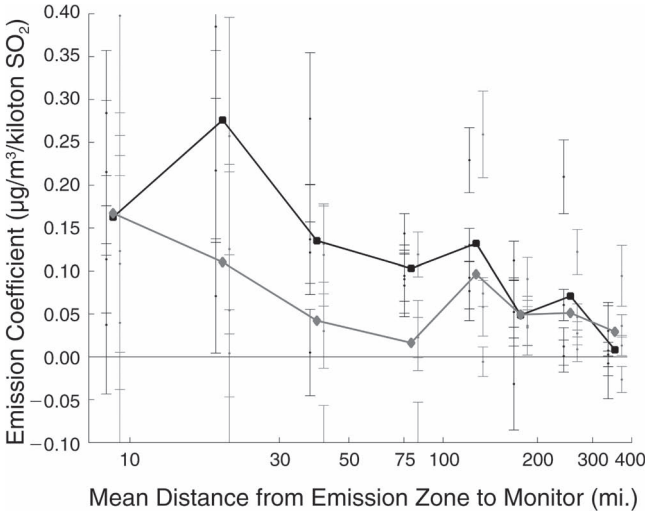


Figure 11. Estimated effects of emissions on ambient $PM_{2.5}$ by distance from source (logarithmic scale), for 116 northern monitors.

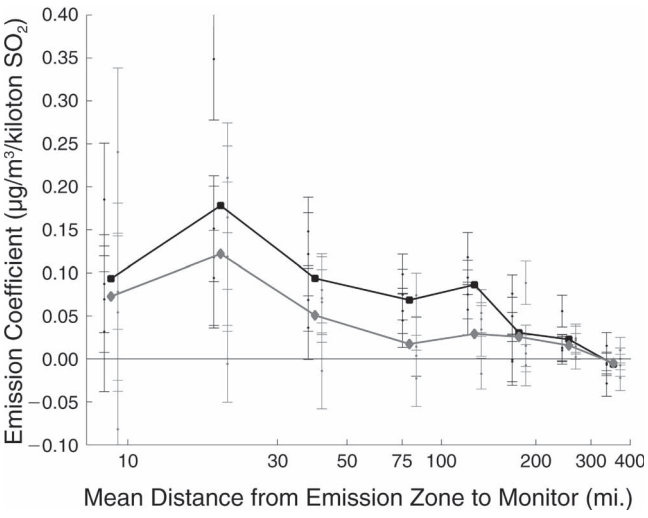


Figure 13. Estimated effects of emissions on ambient $PM_{2.5}$ by distance from source (with monthly fixed effects) (logarithmic scale), for all 193 monitors.

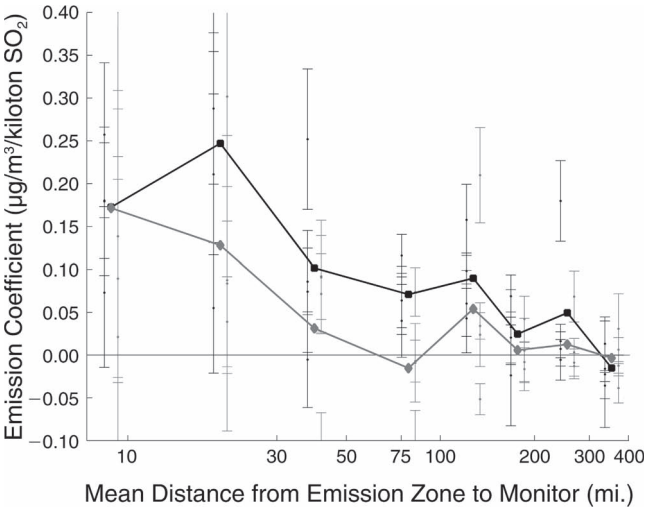


Figure 14. Estimated effects of emissions on ambient $PM_{2.5}$ by distance from source (with monthly fixed effects) (logarithmic scale), for 116 northern monitors.

monitor was quite similar to those shown in Figures 10 to 12. There were, however, two important distinctions. First, the SRCs were consistently smaller when the monthly fixed effects were included, especially for the more distant sources. Second, the SRCs actually turned negative for several of the most distant sources when the monthly fixed effects were included. The question of whether to include monthly variables or not is discussed further in the next section.

Finally, we note the important differences in total emissions in each of the zones used in the estimations. Because there were more power plants in the larger, more distant zones, it was not surprising that average emissions between the closest-in and farthest-out zones varied by two orders of magnitude, as shown in Table 12. Although not relevant to the estimation of the emissions concentrations gradients, these differences are key to the counterfactuals developed below, in the Counterfactual Analysis section.

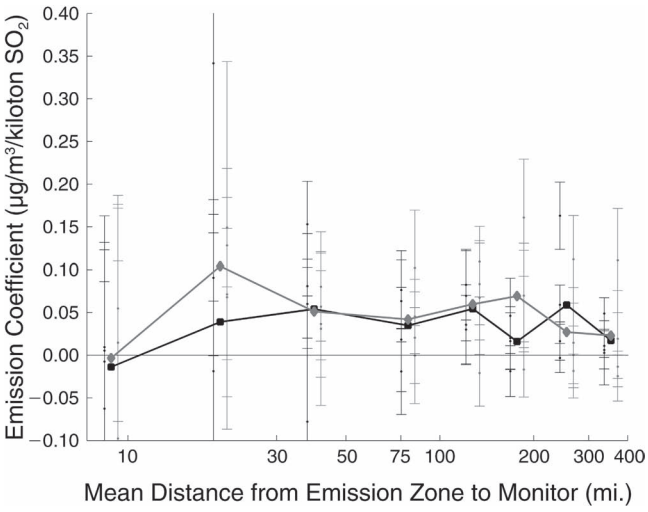


Figure 15. Estimated effects of emissions on ambient $PM_{2.5}$ by distance from source (with monthly fixed effects) (logarithmic scale), for 77 southern monitors.

Choosing Among Alternative Model Specifications

Table 10 shows the results from 42 different models for estimating the relationship between $PM_{2.5}$ concentrations and stationary source emissions. This was a representative set of the total set of model specifications we examined. Of these, we chose only four for the purposes of estimating the effects of the Title IV Phase 2 emission reductions:

- TEMP × EMIS with emission zones to 300 miles, without monthly dummy variables;
- TEMP × EMIS with emission zones to 400 miles, without monthly dummy variables;
- TEMP × EMIS with emission zones to 300 miles, with monthly dummy variables; and
- TEMP × EMIS with emission zones to 400 miles, with monthly dummy variables.

Of these four, our preferred model was the fourth one, at the end of the list. In this section we describe the winnowing process that narrowed our selection to these four, and finally to this one, model.

First, we relied on the average monitor-specific MSE to eliminate the models with TEMP only, models in which temperature and emissions were entered linearly as well as NO_x models and the regional models. Although the southern model had relatively attractive MSEs, the northern model did not, and we could not choose one without choosing the other. As described above, we decided on FGLS over MLE as an estimating strategy because it offered greater convenience (faster run times) without a compensating disadvantage. We eliminated specifications with an emission zone between 400 and 500 miles because these specifications performed no better, and in some cases worse, than those without that zone.

Choosing between models with and without months required us to resolve a dilemma that frequently plagues statistical estimation with variables that are highly correlated. As noted above, failing to include relevant variables in a regression can lead to omitted-variable bias if any of the variables that were included are correlated with the excluded variables. However, including them has pitfalls, too. If the correlations between the two variables are strong enough, then accurate estimation of the coefficients can be very difficult. Sometimes pairs of such variables will be

insignificant, with very large standard errors, or if significant they might have implausibly large coefficients with opposite signs. In the limit, if the variables are perfectly correlated the model cannot be estimated at all unless one or more variables are dropped. Our inability to estimate a model with both SO_2 and NO_x in it is an example.

A set of variables that is nearly dependent linearly can also cause multicollinearity problems even if no pair of them has a particularly high correlation. Sometimes a situation like this can be difficult to spot. Here, it happened that the monthly dummy variables were not very highly correlated with any of the emission variables, even though these seasonal variables were highly correlated with electricity production in total. It was possible we would still have multicollinearity problems even though no pair of variables were obviously correlated with each other. In such situations, one useful tool that has been recommended for identifying multicollinearity is the condition number,

$$\gamma = \left(\frac{\lambda_{max}}{\lambda_{min}}\right)^{1/2},$$

where the lambdas are the maximum and minimum characteristic roots of the moment matrix ($\boldsymbol{X}'\boldsymbol{X}$) of the independent variables. Belsley and colleagues (1980) suggested that multicollinearity problems would be present if the condition number exceeded 20. By this criterion, it appeared that we did not need to worry about multicollinearity. The condition number of the moment matrix of a typical specification (including months, interacted temperature and emissions, and emission zones out to 400 miles) was only 14.7. When months were removed from the specification, the condition number dropped to 3.88. It appeared, then, that there was no reason not to include monthly dummy variables in our specifications. In addition, they were highly significant and had obvious explanatory power even with the other variables in the models.

Even though we rejected the no-month specification because of the likelihood of omitted-variable bias in the estimated coefficients, it might nevertheless be of interest to get an idea of the influence of the inclusion of months in the specification. Figure 16 shows a plot of coefficients from the model specification to 400 miles with monthly dummy variables against an otherwise identical specification without monthly dummy variables. As can be seen, except for two outliers (located at the top and the bottom of the figure), the points in the plot lie just below the 45-degree line and closely track it, indicating that the emission coefficients in the specification without the monthly dummy variables were *larger* than the corresponding coefficients with the monthly dummy variables, and the

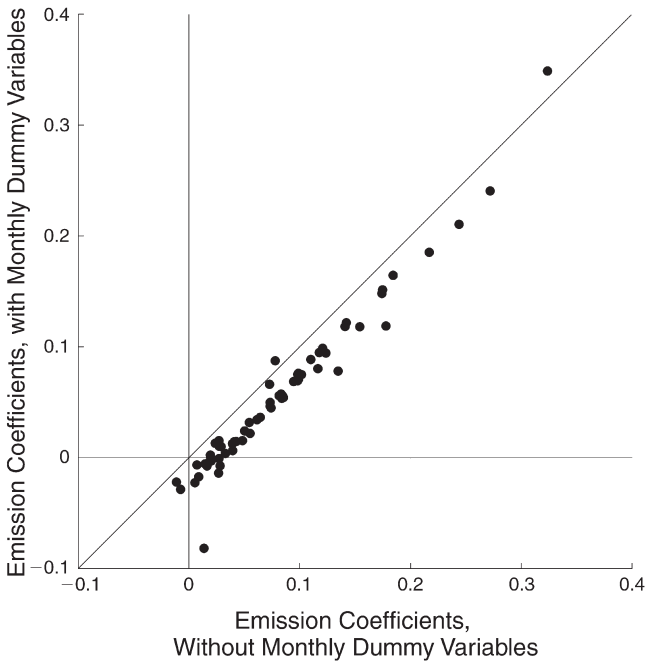


Figure 16. Emission coefficients with monthly dummy variables plotted against emission coefficients without monthly dummy variables.

difference between the two was about the same for all the points except for the outliers. One of the outliers had a difference of about 0.1, and the other had a difference of about 0.025, but with the opposite sign. Overall, the mean difference between the coefficients was 0.026 $\mu g/m^3$ per 1000 tons SO_2. (The specifications used for Figure 16 were the interacted temperature and emissions out to 400 miles. The coefficients were multiplied by the mean temperature for all monitors in the sample.) This is not a small value; as shown below, it has the effect of doubling the predicted effects of emissions on ambient concentrations.

Finally, we chose the specification with the emission zone out to 400 miles instead of the specification without, even though the average MSE of the former was higher than that of the latter (9.37 versus 9.30), because the coefficients as a set had more plausibility. The existence of negative coefficients, implying as it does negative contributions of emissions to $PM_{2.5}$, is an implausible result. (Although negative contributions are implausible, they are not physically impossible, as is well known in the matter of the complex relationships between NO_x and local O_3. Experts at the EPA have suggested to us that something similar might also be going on between long-range transport of SO_2 and ammonia.) Although we were pleasantly surprised to find that considerably fewer than half of our emission coefficients had negative signs in any specification, the number

was never zero. In the specification with the 300-to-400-mile ring, only two of the negative emission coefficients were significantly different from zero; in the specification without this ring, there were five significant negative coefficients. This might have been another instance of omitted-variable bias, and in the absence of any obvious multicollinearity we opted for the more inclusive specification.

MODEL PERFORMANCE

In this section we present the results of a test of our model against emissions and $PM_{2.5}$ data for 2006, an out-of-sample year beyond our basic dataset. This was not a forecast but a test of the model's ability to explain deviations from the mean in a new sample. We found there were 609 $PM_{2.5}$ monitors for which we had sufficient data to obtain average monthly temperature and ambient $PM_{2.5}$ data for every month of 2006. Of these monitors, 164 were also members of the original 1999–2005 dataset used for model estimation. We compared performance on both the 164-monitor sample and the remaining 445 monitors in the 2006 sample, allowing us to extrapolate not only over time, but also to some degree over space. The results are shown in Table 13.

As we did in evaluating the fit of our regression models to the original data, we constructed as a baseline the average squared deviation $V = (1/N)V_i$, where

$$V_i = \frac{1}{12}\sum_{m=1}^{12}(A_{im} - \bar{A}_{i.})^2$$

Table 13. Application of Model Estimates for $PM_{2.5}$ to 2006 Data

	Monitors in Original Dataset	Monitors Not in Original Dataset
Number of monitors	164	445
Mean monthly variance from 2006 annual mean	11.06	9.52
Errors of model estimates		
Mean (MSE)	7.85	7.19
Minimum	1.50	1.03
Maximum	17.7	27.8
Mean reduction in MSE (%)	29.0	24.5
Fractional error (%)		
Model	17.6	18.3
Site means	20.1	20.3

is the variance of monthly means at each monitor. As shown, $V = 11.06$ for the monitors in the original dataset. Similarly, we calculated the model fit as the average deviation or MSE of the monthly predictions from the mean:

$$\mathrm{MSE}_i = \frac{1}{12}\sum_{m=1}^{12}(A_{im} - \hat{A}_{i.})^2.$$

The MSE for all 164 monitors in the original dataset was 7.85; this was 29% smaller than the mean monthly variance for these monitors. Performing the same calculations for the 445 monitors not in the original sample, we found that the average MSE was 7.19. The fact that the MSE in the 2006 sample was smaller than that of the original sample apparently resulted from the smaller variance in the 2006 sample. The MSE for the 2006 sample was 24.5% smaller than the mean monthly variance for these monitors. That is, for the 164 monitors for which we had data for both the original sample years (1999–2005) and the single out-of-sample year (2006), our model explained 29% of the mean deviations observed in the out-of-sample year. For the 445 receptors for which we had complete data for the out-of-sample year but not for the earlier period, our model explained 24.5% of the mean deviations.

We also calculated the fractional bias and fractional error, commonly used criteria for testing the daily performance of air quality models (U.S. EPA 2005). These criteria are defined as follows:

$$\mathrm{FBIAS} = \frac{2}{N}\sum_{i,m}\left(\frac{\hat{A}_{im} - A_{im}}{\hat{A}_{im} + A_{im}}\right);$$

$$\mathrm{FERROR} = \frac{2}{N}\sum_{i,m}\frac{|\hat{A}_{im} - A_{im}|}{\hat{A}_{im} + A_{im}}.$$

Not surprisingly, we found the fractional bias to be quite low (2.3%), reflecting our focus on deviations from the monthly mean rather than daily deviations. As shown in Table 13, the corresponding fractional errors of the site-specific means were 20.1% and 20.3%. By comparison, the fractional errors for the original and 2006 samples were 17.6% and 18.3%, respectively, reflecting a slight improvement of the model versus the site-specific means.

COUNTERFACTUAL ANALYSIS

In this section we apply the regression analysis described above to investigate the effect of power plant emission changes between 1999 and 2005 on ambient $PM_{2.5}$ concentrations.

To perform these calculations we needed a counterfactual, that is, an estimate of the emissions that would have taken place in the absence of Title IV Phase 2 of the 1990 Clean Air Act Amendments (the principal influence of

national environmental policy on power plant emissions during this period). Constructing an appropriate counterfactual is clearly not a strictly empirical matter, and different observers might reasonably disagree about what would have happened if a given policy had not been implemented. It is therefore important to be clear what assumptions were being made about our no-policy alternative.

We began by examining the average monthly $PM_{2.5}$ concentrations for 1999–2005 at the 193 monitoring sites we used for our analyses (Figure 17). The monthly variations were so large that it was difficult to visually discern any obvious trend in the data; for this reason, we also plotted the 12-month moving average in the figure. The moving average began in mid-1999 at 15.4 μg/m^3, reached a low of 12.9 μg/m^3 in May 2004, and increased thereafter to reach 14.4 μg/m^3 by mid-2005.

Especially in light of the historical record, we adopted what we thought was the simplest and most transparent approach to estimating our counterfactual. We assumed that the temporal and spatial pattern of electricity production observed during the period would not have been affected by the emission reductions. In the counterfactual, then, each power plant would have produced exactly the same amount of electricity that it was actually observed to produce during the period. We also assumed that its emission rate, in terms of emissions per million BTU input, would have been unchanged from what was last observed before 2000; in accordance with this assumption, we took the average emission rate for each plant in 1999 and applied it through 2005. These assumptions resulted in an estimate for total SO_2 emissions of 87 million tons for the 1999–2005 period, compared with the 72 million tons actually emitted.

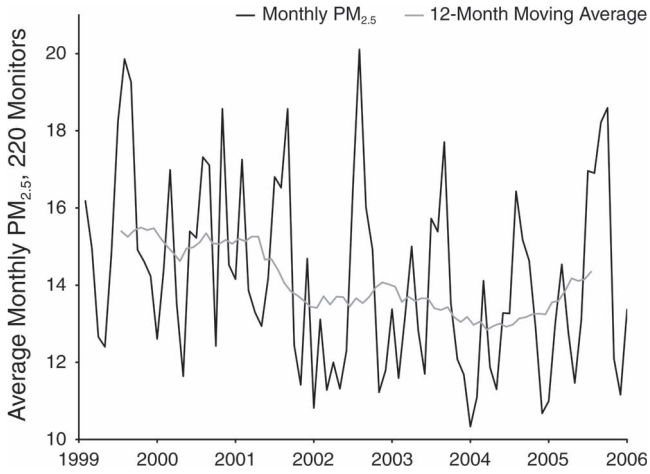


Figure 17. Average monthly (*black line*) and moving average (*gray line*) $PM_{2.5}$ concentrations.

For the counterfactual analysis, we used the model with site–month fixed effects and the interaction of emissions and temperature as independent variables (model 3). Model estimation was by way of FGLS. As an alternative, we used the same model with the inclusion of monthly fixed effects. We used the coefficients estimated in these models to predict what $PM_{2.5}$ concentrations would have been at each site and in each month in the years 2000–2005 if actual source-specific emissions had remained at their 1999 levels. We attributed the difference between these two estimates of $PM_{2.5}$ concentrations to the effects of Title IV Phase 2. The results of this analysis are shown in Table 14. Column one shows that without the Title IV Phase 2 reductions $PM_{2.5}$ concentrations would have been higher by 1.07 to 2.50 μg/m^3 in 2005. The chief cause of the differences was the inclusion or exclusion of the monthly fixed-effects variables in the models. Without monthly fixed effects, Title IV Phase 2 was estimated to have reduced $PM_{2.5}$ concentrations by 2.31 to 2.50 μg/m^3. With the monthly fixed-effects specification, it was estimated to have reduced $PM_{2.5}$ concentrations by 1.07 to 1.17 μg/m^3; on a population-weighted basis applied to all eastern counties, the comparable estimates were 0.89 to 0.94 μg/m^3.

As noted, our preferred specification went out to 400 miles from the monitors. Our preferred unweighted estimate was thus 1.07 μg/m^3, and our preferred population-weighted estimate was 0.89 μg/m^3.

Because observed $PM_{2.5}$ concentrations ranged from 13.0 to 15.9 μg/m^3 during the 2000–2005 period, we estimated that they would have been about 8% to 15% higher without Title IV Phase 2. Although we did not explicitly estimate health effects associated with such changes, they could clearly have been substantial (Bell et al. 2004; Pope and Dockery 2006).

Estimates of the contributions from each ring provided some insights that could be used to evaluate the counterfactual estimate. Recall that most of the regression results depicted in Figures 10 to 15 showed a distinctive and robust pattern with respect to emissions. As distance increased from the monitor, the emission coefficients first increased and then declined. However, when we used these coefficients to project changes in ambient concentrations, we found that the contribution to $PM_{2.5}$ reductions attributable to each zone tended to increase with distance from the zone to the monitor, at least up to a distance of about 300 miles. The reason, of course, was that the outer zones were larger in area, had more power plants in them, and hence had greater total emissions, as shown in Table 12.

Table 14 breaks out the changes in $PM_{2.5}$ concentrations attributable to emissions from each ring, from the

Table 14. Estimated Contribution (in μg/m^3) of Phase 2 Electric Utility Emission Reductions to Changes in Ambient $PM_{2.5}$ Concentrations, by Zone

	Mean 2005 $PM_{2.5}$ Reduction	Zone[a]							
		0–12.5	12.5–25	25–50	50–100	100–150	150–200	200–300	300–400
County Results, Weighted by Population[b]									
400nomo	2.25	0.025	0.042	0.080	0.200	0.452	0.327	0.793	0.332
300nomo	2.03	0.027	0.046	0.089	0.224	0.494	0.345	0.806	—
400mo	0.89	0.018	0.034	0.053	0.116	0.293	0.138	0.311	−0.071
300mo	0.94	0.017	0.036	0.051	0.109	0.283	0.121	0.326	—
Results at 193 Monitors[b]									
	Mean	SD							
400nomo	2.50	0.08	0.045	0.042	0.099	0.290	0.418	0.392	0.904	0.316
300nomo	2.31	0.06	0.051	0.050	0.11	0.33	0.464	0.406	0.901	—
400mo	1.07	0.11	0.036	0.029	0.062	0.186	0.281	0.210	0.405	−0.14
300mo	1.17	0.09	0.035	0.034	0.059	0.179	0.274	0.188	0.404	—

[a] Inner and outer distance from source zone to monitor, in miles.
[b] "nomo" indicates no monthly fixed effects included in model; "mo" indicates inclusion of monthly fixed effects.

innermost ring on the left to the outermost on the right. The values shown are the product of the actual change in emissions in each zone and the regression coefficients for that zone. Specifically, the first row of the table shows the contributions of sources at distances of 0 to 400 miles from the 193 receptors examined to the estimated 2005 $PM_{2.5}$ changes induced by the Title IV Phase 2 policy for the specification without monthly fixed effects. As shown, these values ranged from 0.042 μg/m^3 for units located 12.5–25 miles away to 0.904 μg/m^3 for units located 200–300 miles away. For the model specification with monthly fixed effects, the contributions by zone were uniformly lower. Reflecting the pattern of differences in the SRCs shown in Figures 10 to 15, the major differences in the specifications with and without monthly fixed effects occurred in the outer zones. As can be seen, there was actually a negative value in the 300-to-400-mile zone for the specification with monthly fixed effects. Although we did not necessarily believe in the reality of negative emission coefficients, to avoid bias we left them unchanged for estimating ambient changes.

Parameter Uncertainty Versus Model Uncertainty

Analysts and model builders often make a useful distinction between parameter uncertainty and model uncertainty in assessing the performance of models. The section above discussed the effects of model uncertainty on our results, or at least some of the effects (inasmuch as we only compared a few specifications of the one particular type of model we might have chosen).

For any choice of model there is also parameter uncertainty, that is, the fact that regression coefficients are themselves random variables. The individual coefficients and their standard errors for our top four models are reported in Appendix A. But it is also useful to propagate the errors in the estimated coefficients through emission projections in order to determine the effects of parameter uncertainty on our counterfactual analyses. Accordingly, Table 14 also shows the estimated SDs of the projections of $PM_{2.5}$ improvements for each of the specifications. These SDs were quite small, and we concluded that, as is very often the case, our model uncertainty was much greater than our parameter uncertainty.

Comparison with an Air Quality Model

An additional approach to evaluation of the results would involve a side-by-side comparison of the statistical model results with those of an established air quality model for the same emission reduction and time period, that is, a comparison of the strength of the aggregate source–receptor relationship in the statistical and air quality models. Unfortunately, we were unable to obtain a directly comparable model run; we relied instead on a recently published Regulatory Impact Analysis for CAIR (U.S. EPA 2005).

The Regulatory Impact Analysis relied on the Community Multiscale Air Quality model, a sophisticated air-quality fate and transport model frequently used to model emissions and air quality and to predict changes in air quality associated with anticipated reductions in utility emissions of SO_2 and NO_x for the years 2010 and 2015. The EPA estimated that, as a result of the CAIR-driven emission reductions, $PM_{2.5}$ concentrations would be reduced by 0.73 $\mu g/m^3$ in 2010 and 0.89 $\mu g/m^3$ in 2015 as the average across all grid cells in the eastern United States. On a population-weighted basis, the comparable numbers were about 30% higher (0.96 $\mu g/m^3$ and 1.15 $\mu g/m^3$ for 2010 and 2015, respectively). The SO_2 reductions needed to achieve these reductions were 3.61 million tons in 2010 and 3.97 million tons in 2015. Normalized to the CAIR SO_2 reductions, the population-weighted rates were 0.266 $\mu g/m^3$ per million tons in 2010 and 0.290 $\mu g/m^3$ in 2015.

Turning to the results of our preferred specification, we estimated that Title IV Phase 2 reduced $PM_{2.5}$ concentrations by 0.89 $\mu g/m^3$ in 2005 (Table 14). Normalizing by the Phase 2 SO_2 reductions of 2.34 million tons, the estimated $PM_{2.5}$ sensitivity was 0.38 $\mu g/m^3$ in 2005 — somewhat higher than the sensitivity used in CAIR.

Large differences in methods and context make these estimates difficult to compare directly. Both estimates included the effects of NO_x as well as SO_2, even though our models used only SO_2. Accordingly, it is important to take account of any differences in the ratio of SO_2 to NO_x emissions when comparing CAIR with the Phase 2 reductions. The modeled CAIR analysis assumed 1.25 million tons of NO_x reductions in 2010 and 1.54 million tons in 2015 — less than half as much NO_x on a proportional basis as was achieved under Title IV Phase 2. One would thus have expected our estimate to be higher. Also, during Phase 2, concentrations of $PM_{2.5}$ were 50% higher than the CAIR projections for 2010 and 2015. Perhaps $PM_{2.5}$ formation is less sensitive to stationary source emissions at lower baseline concentrations. This consideration would also tend to raise our estimate relative to the CAIR estimate. And finally there was the assumption about the constancy of the temporal and spatial patterns of electricity production that might not have been comparable between the two analyses. Given these uncertainties, we concluded that our estimates were roughly comparable to those developed for CAIR based on the Community Multiscale Air Quality model, given the fractional bias and fractional errors of the various models.

CONCLUSIONS

In our effort to use statistical modeling to assess what portion, if any, of the observed 1999–2006 changes in ambient $PM_{2.5}$ concentrations can be attributed to Title IV Phase 2 of the 1990 Clean Air Act Amendments, we drew three broad conclusions:

[1] Based on our analysis, it is feasible to use relatively transparent statistical models to conduct an accountability analysis of policies that reduced the long-range transport of power plant emissions, as occurred under Title IV Phase 2. Our coefficients relating utility emissions to ambient $PM_{2.5}$ concentrations were often statistically significant and also passed several tests of plausibility:

- They were overwhelmingly positive, even though their signs were not predetermined by the statistical framework.
- They were sensitive to the distance and direction separating sources and monitors in plausible ways.
- The sensitivity of PM to emission sources increased up to a distance of 12.5 to 25 miles between source and monitor and thereafter declined, becoming statistically insignificant and unstable at distances of about 300 to 400 miles.
- The coefficients also tended to be higher for sources to the west of the monitor than for sources to the east.
- Originally estimated on monthly data for 193 monitors over the period 1999–2005, our preferred model performed equally well for the same 193 monitors in 2006 as well as for an additional 217 monitors not in the original monitor sample in 2006.

[2] Although we found substantial model uncertainty, our preferred specification suggested that Title IV Phase 2 had reduced $PM_{2.5}$ concentrations by an average of 1.07 $\mu g/m^3$, with an SD of 0.11 $\mu g/m^3$, compared with our counterfactual (defined as there having been no change in emission rates per unit of energy input [1 million BTUs]), over the time period. On a population-weighted basis, the comparable reduction in $PM_{2.5}$ was 0.89 $\mu g/m^3$.

[3] Compared with the air quality fate and transport modeling used by the EPA to estimate air quality improvements associated with CAIR for 2010 and 2015 (when baseline $PM_{2.5}$ concentrations were expected to be about one-third lower), our statistical model yielded results that were roughly comparable per ton of SO_2 reduced and well within the estimated confidence intervals of the model.

Based on this study, we believe there are a number of fruitful directions for future research. Specifically:

- Because several years have passed since 2005, the year of the most recent data used in our study, and because we have now developed the methods and computational infrastructure necessary to perform the analysis,

the framework we used here could be extended to newer emissions and air quality data to see if our results can be corroborated.

- Although our focus on the air quality impacts of pollution reduction policies is certainly important, we believe it is even more important to estimate the mortality reductions and other human health effects associated with such policies.

REFERENCES

Bell ML, Davis DL, Gouveia N, Borja-Aburto VH, Cifuentes LA. 2006. The avoidable health effects of air pollution in three Latin American cities: Santiago, Sao Paulo, and Mexico City. Environ Res 100:431–440.

Bell ML, Samet JM, Dominici F. 2004. Time-series studies of particulate matter. Annu Rev Public Health 25:247–280.

Belsley DW, Kuh E, Welsch RE. 1980. Regression Diagnostics. John Wiley & Sons, New York, NY.

Breitner S, Stölzel M, Cyrys J, Pitz M, Wölke G, Kreyling W, Küchenhoff H, Heinrich J, Wichman HE, Peters A. 2009. Short-term mortality rates during a decade of improved air quality in Erfurt, Germany. Environ Health Perspect 117: 448–454.

Chay KY, Dobkin C, Greenstone M. 2003. The Clean Air Act of 1970 and adult mortality. J Risk Uncert 27:279–300.

Chay KY, Greenstone M. 2003. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. Quarter J Econ 118:1121–1167.

Clancy L, Goodman P, Sinclair H, Dockery DW. 2002. Effect of air pollution control on death rates in Dublin, Ireland: An intervention study. Lancet 360:1210–1214.

Committee on Research Priorities for Airborne Particulate Matter, National Research Council. 1998. Research Priorities for Airborne Particulate Matter: No. 1. Immediate Priorities and a Long-Range Research Portfolio. National Academy Press, Washington, DC.

Dominici F, Peng RD, Zeger SL, White RH, Samet JM. 2007. Particulate air pollution and mortality in the United States: Did the risks change from 1987 to 2000? Am J Epidemiol 166:880–888.

Friedman MS, Powell KE, Hutwagner L, Graham LM, Teague WG. 2001. Impact of changes in transportation and commuting behaviors during the 1996 Summer Olympic Games in Atlanta on air quality and childhood asthma. J Amer Med Assoc 285:897–905.

Greene WH. 1993. Econometric Analysis. 2nd Edition. Prentice-Hall, New York, NY.

Health Accountability Working Group. 2003. Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research. Communication 11. Health Effects Institute, Boston, MA.

Hedley AJ, McGhee SM, Barron B, Chau P, Chau J, Thach TQ, Wong T-W, Loh C, Wong C-M. 2008. Air pollution: Costs and paths to a solution in Hong Kong—understanding the connections among visibility, air pollution, and health costs in pursuit of accountability, environmental justice, and health protection. J Toxicol Environ Health A 71: 544–554.

Hsiao C. 1986. Analysis of Panel Data. Econometric Society monograph 11. Cambridge University Press, New York, NY.

Hubbell BJ, Hallberg A, McCubbin DR, Post E. 2006. Health-related benefits of attaining the 8-hr ozone standard. Environ Health Perspect 113:73–82.

Krupnick A, Morgenstern R. 2002. The future of benefit-cost analyses of the Clean Air Act. Annu Rev Public Health 23:427–448.

Lee JT, Son JY, Cho YS. 2007. Benefits of mitigated ambient air quality due to transportation control on childhood asthma hospitalization during the 2002 summer Asian Games in Busan, Korea. J Air Waste Manag Assoc 57:968–973.

Li Y, Wang W, Kan H, Xu X, Chen B. 2009. Air quality and outpatient visits for asthma in adults during the 2008 Summer Olympic Games in Beijing. Sci Total Environ 408:1226–1227.

Lin C. 2010. A spatial econometric approach to measuring air pollution externalities: An application to ozone smog. J Reg Anal Pol 40:1–19.

Pope CA III, Dockery DW. 2006. Health effects of fine particulate air pollution: Lines that connect. J Air Waste Manag Assoc 56:709–742.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000. National Morbidity, Mortality, and Air Pollution Study. Health Effects Institute, Cambridge, MA.

U.S. Environmental Protection Agency. 1997. The Benefits and Costs of the Clean Air Act, 1970 to 1990. Office of Air and Radiation/Office of Policy, Washington, DC. Available from *www.epa.gov/oar/sect812/copy.html*. Accessed 10/2011.

U.S. Environmental Protection Agency. 1999. The Benefits and Costs of the Clean Air Act, 1990 to 2010. Office of Air and Radiation/Office of Policy, Washington, DC. Available from *www.epa.gov/oar/sect812/prospective1.html*. Accessed 10/2011.

U.S. Environmental Protection Agency. 2005. Technical Support Document for the Final Clean Air Interstate Rule: Air Quality Modeling. Office of Air Quality Planning and Standards, Research Triangle Park, NC. Available from *www.epa.gov/cair/pdfs/finaltech02.pdf*. Accessed 1/2012.

Wu S, Deng F, Niu J, Huang Q, Liu Y, Guo X. 2010. Association of heart rate variability in taxi drivers with marked changes in particulate air pollution in Beijing in 2008. Environ Health Perspect 118:87–91.

APPENDIX A. Coefficients and Standard Errors of Principal Models

Table A.1 shows the coefficients and standard errors of the four models described in the Model Results section of the report. The variable names in the left-hand column are defined as follows: temp is the mean monthly temperature, in Kelvin, at each of the 193 monitor sites; tempdev is the deviation in temperature from the monthly mean; and the variables feb through dec are the dummy variables for month. The remaining rows of the table give the values for the emission coefficients, which are coded for radius (R) and direction (D). The numbers 12.5, 25, 50, 100, 150, 200, 300, and 400 are the radius (in miles) of the outer boundary of each zone (the inner radius being of course the outer radius of the next-smaller zone). The numbers 1 through 8 indicate the directional octant, counting counterclockwise from the 12 o'clock position: the four zones in each ring to the west of the monitor are numbered 1 through 4; to the east, 5 through 8. See text for more information. Coefficients marked * or ** were significantly different from zero at the 5% and 1% confidence levels, respectively.

Table A.1. Coefficients and Standard Errors of Principal Models

	Model Specifications							
	Without Monthly Dummy Variables				With Monthly Dummy Variables			
	Max. Zone = 400 Miles Mean MSE = 9.37		Max. Zone = 300 Miles Mean MSE = 9.3		Max. Zone = 400 Miles Mean MSE = 9.46		Max. Zone = 300 Miles Mean MSE = 9.54	
Variable Name	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
temp	0.045**	0.005	0.050	0.004				
tempdev					0.233**	0.021	0.169**	0.014
feb					0.608**	0.165	0.817**	0.112
mar					−1.345**	0.164	−0.801**	0.112
apr					−1.845**	0.175	−0.878**	0.117
may					−1.311**	0.166	−1.292**	0.113
jun					0.349*	0.163	0.210	0.111
jul					2.810**	0.171	2.194**	0.115
aug					1.531**	0.178	0.631**	0.119
sep					−0.011	0.161	0.448**	0.110
oct					−1.060**	0.164	−1.601**	0.112
nov					−1.954**	0.173	−1.768**	0.116
dec					−1.096**	0.155	−0.932**	0.110
R12.5D1	0.217**	0.030	0.229**	0.031	0.185**	0.033	0.230**	0.029
R12.5D2	0.054	0.033	0.065*	0.033	0.032	0.035	0.066*	0.030
R12.5D3	0.098**	0.030	0.118**	0.030	0.069**	0.031	0.107**	0.029
R12.5D4	0.078**	0.028	0.091**	0.028	0.087**	0.028	0.083**	0.025
R12.5D5	0.085*	0.042	0.092*	0.043	0.054	0.046	0.118**	0.040
R12.5D6	0.135**	0.047	0.120*	0.048	0.078	0.052	0.102**	0.049
R12.5D7	0.013	0.055	−0.003	0.054	−0.082	0.058	−0.025	0.050
R12.5D8	0.272**	0.047	0.293**	0.048	0.241**	0.049	0.243**	0.040

(*Table continues on next page*)

Table A.1 (*Continued*). Coefficients and Standard Errors of Principal Models

	Model Specifications							
	Without Monthly Dummy Variables				With Monthly Dummy Variables			
	Max. Zone = 400 Miles Mean MSE = 9.37		Max. Zone = 300 Miles Mean MSE = 9.3		Max. Zone = 400 Miles Mean MSE = 9.46		Max. Zone = 300 Miles Mean MSE = 9.54	
Variable Name	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
R25D1	0.124**	0.026	0.136**	0.026	0.094**	0.028	0.073**	0.025
R25D2	0.175**	0.029	0.189**	0.029	0.151**	0.031	0.203**	0.028
R25D3	0.324**	0.034	0.321**	0.034	0.349**	0.036	0.350**	0.031
R25D4	0.141**	0.039	0.147**	0.039	0.118**	0.041	0.166**	0.035
R25D5	0.178**	0.041	0.168**	0.041	0.119**	0.043	0.164**	0.038
R25D6	0.244**	0.029	0.255**	0.029	0.210**	0.032	0.221**	0.031
R25D7	0.184**	0.038	0.193**	0.038	0.164**	0.042	0.259**	0.036
R25D8	0.015	0.022	0.032**	0.022	−0.006	0.022	0.044*	0.019
R50D1	0.142**	0.023	0.161**	0.023	0.122**	0.024	0.152**	0.023
R50D2	0.064**	0.017	0.072**	0.017	0.036*	0.018	0.050**	0.017
R50D3	0.174**	0.019	0.190**	0.019	0.148**	0.020	0.169**	0.019
R50D4	0.094**	0.018	0.097**	0.018	0.069**	0.018	0.108**	0.016
R50D5	0.027	0.021	0.036	0.021	−0.014	0.022	−0.019	0.020
R50D6	0.116**	0.018	0.123**	0.018	0.080**	0.019	0.065**	0.017
R50D7	0.099**	0.024	0.106**	0.025	0.071**	0.026	0.061**	0.023
R50D8	0.073**	0.018	0.080**	0.018	0.066**	0.019	0.052**	0.017
R100D1	0.102**	0.014	0.110**	0.014	0.075**	0.015	0.069**	0.012
R100D2	0.074**	0.015	0.089**	0.015	0.045*	0.016	0.086**	0.014
R100D3	0.081**	0.013	0.088**	0.013	0.056**	0.013	0.094**	0.012
R100D4	0.121**	0.011	0.128**	0.011	0.099**	0.012	0.129**	0.010
R100D5	0.033**	0.012	0.045**	0.011	0.004	0.012	−0.022*	0.010
R100D6	0.099**	0.012	0.110**	0.012	0.074**	0.013	0.031**	0.012
R100D7	0.041*	0.017	0.055**	0.016	0.014	0.017	−0.049**	0.016
R100D8	0.005	0.015	0.011	0.015	−0.023	0.016	−0.042*	0.015
R150D1	0.083**	0.014	0.093**	0.014	0.057**	0.015	0.092**	0.012
R150D2	0.154**	0.014	0.164**	0.014	0.118**	0.014	0.166**	0.013
R150D3	0.098**	0.013	0.107**	0.013	0.075**	0.014	0.130**	0.013
R150D4	0.118**	0.010	0.126**	0.009	0.095**	0.010	0.138**	0.008
R150D5	0.073**	0.008	0.082**	0.008	0.046**	0.008	0.047**	0.007
R150D6	0.009	0.009	0.018*	0.008	−0.017	0.009	−0.038**	0.007
R150D7	0.084**	0.013	0.093**	0.013	0.054**	0.014	0.022	0.012
R150D8	0.061**	0.015	0.069**	0.015	0.034*	0.016	0.082**	0.014
R200D1	0.027	0.014	0.028*	0.014	−0.001	0.015	0.022	0.013
R200D2	0.020	0.011	0.018	0.011	−0.003	0.012	0.014	0.011
R200D3	0.073**	0.010	0.077	0.010	0.050**	0.011	0.100**	0.010
R200D4	0.099**	0.010	0.101**	0.010	0.076**	0.011	0.090**	0.009
R200D5	0.110**	0.012	0.110**	0.012	0.088**	0.013	0.139**	0.010
R200D6	0.048**	0.011	0.053**	0.011	0.015	0.012	0.001	0.010
R200D7	0.039**	0.010	0.043**	0.010	0.006	0.011	−0.002	0.009
R200D8	0.028**	0.011	0.036**	0.011	−0.007	0.012	0.018	0.011

(*Table continues on next page*)

Table A.1 (*Continued*). Coefficients and Standard Errors of Principal Models

	Model Specifications							
	Without Monthly Dummy Variables				With Monthly Dummy Variables			
	Max. Zone = 400 Miles Mean MSE = 9.37		Max. Zone = 300 Miles Mean MSE = 9.3		Max. Zone = 400 Miles Mean MSE = 9.46		Max. Zone = 300 Miles Mean MSE = 9.54	
Variable Name	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
R300D1	0.039**	0.007	0.039**	0.007	0.013	0.008	0.033**	0.007
R300D2	0.029**	0.008	0.028**	0.007	0.010	0.008	0.035**	0.007
R300D3	0.024**	0.007	0.022**	0.007	0.013	0.008	0.044**	0.007
R300D4	0.085**	0.009	0.089**	0.009	0.056**	0.009	0.063**	0.008
R300D5	0.050**	0.008	0.046**	0.008	0.024**	0.008	0.065**	0.006
R300D6	0.043**	0.007	0.041**	0.007	0.015	0.008	0.013*	0.006
R300D7	0.019**	0.007	0.022**	0.007	0.002	0.007	−0.014*	0.006
R300D8	0.055**	0.008	0.053**	0.007	0.022*	0.008	0.039**	0.007
R400D1	0.017**	0.006			−0.005	0.006		
R400D2	0.027**	0.007			0.015	0.008		
R400D3	0.007	0.006			−0.006	0.006		
R400D4	−0.008	0.007			−0.029**	0.007		
R400D5	0.027**	0.007			0.010	0.007		
R400D6	0.020**	0.006			0.000	0.006		
R400D7	−0.011	0.007			−0.022**	0.007		
R400D8	0.016**	0.006			−0.007	0.006		

ABOUT THE AUTHORS

Richard D. Morgenstern is a senior fellow at Resources for the Future, in Washington, D.C. His research focuses on the economic analysis of environmental issues with an emphasis on the costs, benefits, evaluation, and design of environmental policies, especially economic incentive measures. His work also focuses on climate change, including the design of cost-effective policies to reduce emissions in the United States and abroad. Previously he served at the EPA as deputy administrator (1993); assistant administrator for policy, planning, and evaluation (1991–1993); and director of the Office of Policy Analysis (1983–1995). Formerly a tenured professor at the City University of New York, he has taught recently at Oberlin College, the Wharton School of the University of Pennsylvania, Yeshiva University, and Howard University. He has also served on expert committees of the National Academy of Sciences and as a consultant to various organizations. He holds a Ph.D. in economics from the University of Michigan.

Winston Harrington, associate research director and a senior fellow at Resources for the Future, in Washington, D.C., focuses his research on urban transportation, motor vehicles and air quality, and problems of estimating the costs of environmental policy. He has worked extensively on the economics of enforcing environmental regulations, the health benefits derived from improved air quality, the costs of waterborne disease outbreaks, endangered-species policy, federal rulemaking procedures, and the economics of outdoor recreation. He has served as a consultant to U.S. state and federal governments, the World Bank, and the Harvard Institute for International Development and has worked in Lithuania, Mexico, and Poland. He also is on the adjunct faculty at Georgetown University. He holds a Ph.D. in city and regional planning from the University of North Carolina at Chapel Hill and an M.A. in mathematics from Cornell University.

Jhih-Shyang Shih is a fellow at Resources for the Future, in Washington, D.C. His research interests lie in quantitative analysis of environmental and resource policy and decision making. He has extensive experience with modeling to study air quality, risk, surface water, and solid waste management and has studied the costs of environmental protection, technology adoption, and renewable

energy. His recent work has focused on O_3 and PM control, recycling, small water systems, and space solar power. His articles have appeared in such publications as the *Review of Economics and Statistics*, the *Journal of Hazardous Materials*, the *Journal of Water Resources Planning and Management*, and the *European Journal of Operational Research*. He has received fellowships from the American Association for the Advancement of Science, the EPA, and Carnegie Mellon University. He holds a Ph.D. in system analysis and economics for public decision making from Johns Hopkins University and an M.S. in environmental engineering from National Cheng-Kung University, in Tainan City, Taiwan.

Michelle L. Bell is an associate professor of environmental health at the Yale University School of Forestry and Environmental Studies. Her research addresses air pollution and human health by integrating several disciplines, including environmental engineering and epidemiology. The overall aim of her work is to answer scientific questions about how air pollution affects health and to perform policy-relevant research that contributes to well-informed decision making and greater public understanding of environmental health hazards. She is also interested in the health effects of climate change, in particular the potential changes in air pollution caused by climatic change and the subsequent health impacts. She holds a Ph.D. in environmental engineering and an M.S.E. in environmental management from Johns Hopkins University as well as an M.S. in environmental engineering from Stanford University.

ABBREVIATIONS AND OTHER TERMS

BTU	British thermal unit
CAIR	Clean Air Interstate Rule
CAM	Clean Air Markets
EPA	Environmental Protection Agency
FGLS	feasible generalized least squares
ML	maximum likelihood
MLE	maximum likelihood estimation
MSE	mean squared error
NAAQS	National Ambient Air Quality Standards
NEI	National Emissions Inventory
NO_2	nitrogen dioxide
NO_x	nitrogen oxides
O_3	ozone
OLS	ordinary least squares
PM	particulate matter
$PM_{2.5}$	particulate matter $\leq$ 2.5 μg in aerodynamic diameter
PM_{10}	particulate matter $\leq$ 10 μg in aerodynamic diameter
RFA	request for applications
SIC	Standard Industrial Classification
SO_2	sulfur dioxide
SRC	source–receptor coefficient

Research Report 168, *Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments*, Morgenstern et al.

INTRODUCTION

PROGRAM BACKGROUND

Over the past two decades, scientific studies have found that long-term exposure to particulate matter (PM) ≤ 2.5 µm in aerodynamic diameter ($PM_{2.5}$*) is associated with premature mortality and adverse health outcomes. In the 1990s, studies of large cohorts, such as the American Cancer Society Cancer Prevention Study II and Harvard Six Cities studies, reported associations between premature cardiopulmonary mortality and long-term $PM_{2.5}$ exposure. An HEI-led reanalysis of the data from these two studies confirmed the original findings, and further analyses with additional years of follow-up have strengthened findings that exposure to $PM_{2.5}$ is associated with premature mortality from cardiovascular disease (Health Effects Institute 2000a). In a recent analysis of the Harvard Six Cities Cohort data, Lepeule and colleagues (2012) reported a 14% increase in the risk of all-cause premature mortality and a 24% increase in the risk of premature cardiovascular mortality for each 10-µg/m^3 increase in long-term $PM_{2.5}$ exposure. In HEI Report 140, published in 2009, Krewski and colleagues used a random-effects model that controlled for personal and ecologic confounders and reported a 24% increase in the risk of premature mortality from ischemic heart disease associated with a 10-µg/m^3 increase in long-term $PM_{2.5}$ exposure.

As understanding of the adverse health effects associated with exposure to $PM_{2.5}$ has increased, ambient $PM_{2.5}$ concentrations have increasingly been regulated in the United States and Europe. The monitoring and regulation of $PM_{2.5}$ by the U.S. Environmental Protection Agency (EPA), beginning with the 1997 revision of the National Ambient Air Quality Standards (NAAQS) for PM, has been credited with improving air quality nationwide, with considerably greater improvements in the most heavily exposed areas of the eastern United States (Krewski et al. 2009). However, it is still not entirely clear whether specific regulatory programs designed to reduce ambient $PM_{2.5}$ concentrations are responsible for improvements in health, such as the increased life expectancy associated with reductions in $PM_{2.5}$ concentrations on a national scale that was reported by Pope and colleagues (2009). It is desirable to verify that air quality regulations have resulted in improved air quality, improved health, and reduced mortality, but evidence for verification — particularly in terms of health outcomes — has not often been systematically and prospectively collected and has been very difficult to establish retrospectively. There has thus been a long-standing need for research to assess the performance of environmental regulatory policy, an effort that has been termed accountability research or, more recently, health outcomes research (Health Effects Institute 2009).

HEI launched an initiative to improve the evidentiary and methodologic bases for assessing the health effects of regulations and other actions or situations resulting in improved air quality, beginning with its 2000–2005 Strategic Plan (Health Effects Institute 2000b). In 2003, HEI published a monograph setting out a conceptual framework to address health outcomes research needs, identifying the types of evidence required as well as the methods by which this evidence could be obtained (HEI Accountability Working Group 2003). HEI funded nine studies through four Requests for Applications (RFAs) issued between 2002 and 2004. Each of these RFAs solicited studies to measure the health impacts of actions taken to improve air quality (see the Preface to this report). The study led by Dr. Richard D. Morgenstern of Resources for the Future was funded under RFA 04-4, *Measuring the Health Impacts of Measures Taken To Improve Air Quality*, which sought proposals for studies of the health effects of real-world situations where there were either planned actions to improve air quality or other situations resulting in marked air quality improvements.

Dr. Richard D. Morgenstern's 3-year study, "Accountability Assessment of Title IV of the Clean Air Act Amendments of 1990," began in March 2006. Total expenditures were $785,599. The draft Investigators' Report from Morgenstern and colleagues was received for review in April 2010. A revised report, received in January 2011, was accepted for publication in February 2011. During the review process, the HEI Health Review Committee and the investigators had the opportunity to exchange comments and to clarify issues in both the Investigators' Report and the Review Committee's Commentary.

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

In response to RFA 04-4, Morgenstern and colleagues initially proposed an analysis of the effects of reductions in pollutant emissions from power plants on $PM_{2.5}$ concentrations in the eastern United States between 1999 and 2005, using a statistical model linking emissions and air quality monitoring data, with an additional time-series analysis of daily all-cause, cardiovascular, and respiratory mortality in selected cities before and after implementation of the emissions reduction programs. Title IV of the Clean Air Act Amendments of 1990, intended to reduce acid rain, a then-growing issue, mandated major reductions in emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x) from coal-fired electric power plants. Reducing SO_2 and NO_x, which react in the atmosphere to form PM, reduces airborne concentrations of $PM_{2.5}$ in turn. At the request of the HEI Research Committee, because of concerns about whether the health outcomes data were adequate to enable the investigators to detect changes in mortality patterns that could be attributed to the changes in air pollutant concentrations, Morgenstern revised his proposal to focus solely on modeling the effects of emissions reductions on $PM_{2.5}$ concentrations.

REGULATORY BACKGROUND

In 1990 the U.S. Congress revised the Clean Air Act with new amendments designed to curb the growing problem of acid rain, among other air quality issues. This legislation also contained provisions for a national permits program and an improved enforcement program as well as for encouraging the use of market-based principles, performance-based standards, emissions banking and trading, the use of low-sulfur coal and sulfur-scrubbing technologies, and energy waste reduction (U.S. EPA 1990).

Title IV of the Clean Air Act of 1990, entitled "Acid Deposition Control," called for a permanent 10-million-ton reduction in SO_2 emissions from 1980 levels, to be accomplished in two phases. The first phase, commencing January 1, 1995, required 110 power plants with relatively high emissions to reduce their emissions to 2.5 lb SO_2 per million British thermal units (BTUs) of their average fuel use in 1985–1987. The second phase, effective January 1, 2000, required approximately 2000 power plants to reduce their emissions to 1.2 lb SO_2 per million BTUs of their average fuel use in 1985–1987. Title IV also required installation of continuous monitoring equipment for SO_2 emissions in both phases to ensure compliance with the law and to track improvements, incentives for early adoption of the standards and modernization of equipment, and penalties for noncompliance. Reductions in NO_x emissions were to be determined by way of EPA regulations no later than mid-1992 for certain boilers and 1997 for all remaining boilers (U.S. EPA 1990). In 2000, electric power generation was responsible for 63% of SO_2 emissions and 22% of NO_x emissions in the United States. SO_2 and NO_x can react with other chemical species in the atmosphere to form PM, including $PM_{2.5}$ (U.S. EPA 2002). Thus, the substantial reductions in SO_2 and NO_x required under Title IV of the Clean Air Act could also reasonably be expected to reduce concentrations of $PM_{2.5}$, particularly in the eastern United States, where coal combustion was a major source of energy for electric power generation and a significant contributor to $PM_{2.5}$.

Beginning with the 1997 NAAQS for PM, the EPA set standards for $PM_{2.5}$ concentrations. As part of implementing this new rule, $128 million was appropriated for the construction of a monitoring network to measure ambient concentrations of $PM_{2.5}$. Referred to in the report by Morgenstern and colleagues as the EPA's Air Quality System monitoring network, this nationwide array of $PM_{2.5}$ monitoring stations had grown to 1022 active sampling sites by the year 2000. The site locations were selected by the EPA and are generally in populous areas of the country where significant 24-hour and annual exposures to $PM_{2.5}$ were expected (U.S. EPA 1998).

SCIENTIFIC BACKGROUND

When evaluating a regulatory action intended to improve air quality, either prospectively or retrospectively, EPA scientists frequently employ a chemical transport model, such as the Community Multiscale Air Quality (CMAQ) modeling system. The CMAQ model or similar chemical transport models can be used either to estimate future concentrations of criteria pollutants based on expected implementation of a regulation that reduces emissions or to estimate changes in ambient concentrations of pollutants based on known reductions in source emissions and known or projected meteorologic and atmospheric conditions. Chemical transport models simulate chemical and physical processes that are believed to play an important role in pollutant formation, transformation, transport, and fate. The CMAQ model has been widely used by U.S. states for statewide NAAQS implementation plans, by the EPA to assess national programs, and by the National Weather Service to produce daily air quality forecasts (U.S. EPA 2012).

One key limitation to the use of chemical transport models, including CMAQ estimates of the effects of regulatory actions, is that they are based on modeled estimates of pollutant concentrations and not on monitoring data. For the current study, Morgenstern and colleagues proposed a model that was based both on data for NO_x and SO_2 emission sources and on $PM_{2.5}$ concentration data from the Air Quality System monitoring network. This innovative approach was data-driven, in that it made use

of measured values of emissions and pollutant concentrations and was inherently observational. The statistical models that the investigators developed to link changes in SO_2 and NO_x emissions to changes in ambient $PM_{2.5}$ concentrations were broadly based on the source–receptor models that are widely used for source apportionment (i.e., analyzing source contributions to observed pollutant concentrations). More specifically, the investigators based their work on a "spatial econometric" approach, incorporating a statistical accounting of emissions in the manner of economic analysis, adapted for the current purposes of associating emissions and air pollution concentrations (Lin 2010). The investigators stated that this was an appropriate approach to an accountability, or intervention, study because it retrospectively assessed the relationships between the implementation of a policy and the resulting measured changes in air quality. It therefore required fewer and less complicated assumptions about the nature and kinetics of complex chemical and atmospheric behavior than the CMAQ modeling system, because the air quality changes were measured directly. However, unlike the chemical transport model approach, the approach of Morgenstern and colleagues can only be applied in a retrospective fashion; without the use of simulated data, it cannot be applied prospectively to forecast concentration changes resulting from proposed regulatory efforts.

This Commentary is intended to aid the sponsors of HEI and the public by highlighting both the strengths and limitations of the study and by placing the Investigators' Report into scientific and regulatory perspective.

APPROACH

The investigators' specific aims for the study were as follows:

1. To assess what portion, if any, of the observed reductions in ambient concentrations of $PM_{2.5}$ that occurred in the United States in the years 1999–2005 could be credited to emissions reductions resulting from the implementation of Title IV Phase 2 of the 1990 Clean Air Act Amendments; and
2. To develop a statistical modeling approach to link observed changes in emissions of SO_2 and NO_x from power plants to changes in $PM_{2.5}$ concentrations.

The overall intention of Morgenstern and his team was to assess the linkages among the regulations reducing emissions of SO_2 and NO_x, the implementation of the regulations, and the changes in air quality measured as ambient $PM_{2.5}$ — all steps in the "chain of accountability" described in the Preface to this report.

METHODS

DATASETS

The models for this study were built using three datasets. Two of these, the EPA's Clean Air Markets (CAM) database and the National Emissions Inventory (NEI) database, provided inventories of source emissions of SO_2 and NO_x. The third, modeled as receptor data, consisted of air quality data from 193 monitors in the EPA's Air Quality System.

CAM Database

The CAM database includes data on power plant emissions obtained from emissions monitoring equipment installed to comply with the Clean Air Act Amendments. It includes sub-hourly readings from continuous emission monitors for all power plants participating in the EPA's Acid Rain Program, the NO_x budget trading program, and other industrial sources participating in the Acid Rain Program. Morgenstern and his team aggregated these readings into a monthly emissions dataset.

NEI Database

The NEI database consists of annual county-level estimates of mobile and area sources, with adjustments for changes in population and the local motor vehicle stock. Because the investigators modeled emissions and air quality levels on a monthly basis, they ultimately chose not to use these data directly and turned to an array of variables to model the same non–power plant emissions and trends. They did, however, use NEI data to cross-check information from the CAM database as well as to keep track of new industrial sources that were opting in to the Acid Rain Program.

EPA's Air Quality System

Data for ambient $PM_{2.5}$ concentrations were obtained through the EPA's Air Quality System. Although some urban monitors measured $PM_{2.5}$ on a daily basis, most were operating one day in three during the study period. This dataset also included monitor location, date and time of sample measurement, and ambient temperature during sampling. As with the emissions data, observations were averaged on a monthly basis.

Data Handling

For emissions, the CAM data were classified by location and date and adjusted for any important NEI database variables (e.g., industrial emissions sources that opted in to the Acid Rain Program during the study period). These

emissions data were then compiled into a monthly emissions dataset. The Air Quality System monitoring dataset was also sorted by date and location, and monthly average values were tabulated for months in which there was an acceptable minimum amount of data (at least two days of data for each of the 84 months of the study period).

Based on the location of the Air Quality System monitors, the investigators defined circular zones of radius 12.5, 25, 50, 100, 150, 200, 300, 400, and 500 miles, according to the location information in the emissions inventory datasets. These zones were divided into wedge-like "sectors" along the principal compass headings and were used to tabulate emissions within them. Monthly average $PM_{2.5}$ readings at the monitors, emissions in the zones surrounding the monitors, meteorologic variables from the Air Quality System database, and a large array of dummy variables and interaction variables to account for unmeasured emissions and factors influencing $PM_{2.5}$ concentrations (such as non-power plant sources and seasonal variations) were combined in a statistical modeling framework relating emissions of SO_2 and NO_x to measured $PM_{2.5}$, as described below.

MODEL BUILDING

Basic Model

The investigators used linear regression models to explore the statistical relationships between source emissions of SO_2 and NO_x, various specifications of dummy variables and interaction variables (referred to in the report as fixed-effects variables), and the monitored concentrations of $PM_{2.5}$. In these models, monthly average monitored concentrations of $PM_{2.5}$ were the "outcome," or "dependent," variable, modeled as a function of power plant emissions of SO_2 and NO_x in the circular zones and of the fixed-effects variables. As described above, the fixed-effects variables were intended to capture area source emissions and site-specific factors that can affect measured $PM_{2.5}$ but were not specifically identified and measured for the study. The investigators also explored various models in which they altered the functional forms (e.g., log transformations) of the emissions variables, fixed-effects variables, and interaction terms to account for possible nonlinear relationships in the model.

The investigators first constructed a relatively simple model in which monthly average $PM_{2.5}$ measured at the monitors was modeled as a function of monthly power plant source emissions in the circular zones, monitor-site temperature (measured as deviation from mean temperature for the study period), and fixed-effects indicator variables for monitor site, year, and month. The investigators controlled for the varying number of monthly values available for the monitors included in the analysis by weighting the data by the number of observations. Because emissions of SO_2 and NO_x were very highly correlated (average within-zone correlation of 0.88), there was some difficulty in fitting the linear regression model when both were included. Consequently, the investigators chose to model the $PM_{2.5}$ concentrations as a function of SO_2 emissions alone, dropping the NO_x emissions from subsequent models.

Model Completion and Selection

Proceeding from this initial model, the investigators built more complex models that included interaction terms combining such factors as temperature and emissions or site and year, exploring the effects of various sizes of zone rings around the monitors (e.g., only up to 200 or 300 miles), weighting emissions in quadrants of the zone rings to account crudely for wind patterns, including fixed-effects variables for season, and so forth. They also discussed and created adjustment schemes for violations of the assumptions of the linear regression model arising from the structure of their emissions and monitoring data, such as spatial autocorrelation (non-independence of observations at nearby monitors and facilities) and heteroskedasticity (differences in patterns of variance in observations across the range of values). Because the standard ordinary least squares model-fitting methods used with linear regression models would not accommodate these corrections, the investigators tested both maximum likelihood and feasible generalized least squares algorithms to fit the models, selecting the latter as the most appropriate for their purposes.

The investigators computed the mean squared error (MSE) of the models — a measure of the net absolute differences between the model-predicted $PM_{2.5}$ concentrations from the data and the actual $PM_{2.5}$ concentrations from observations — as a measure of the fit of the model. (The units for MSE statistics are the units of the measure of interest [in this case $\mu g/m^3$] squared.) The team preferred models where the combination of predictive variables — emissions, fixed effects, and interaction terms — resulted in the lowest MSE.

In Table 11 of the Investigators' Report, the authors presented their various model-fitting results for models with three sizes of emission zones and 14 types of fixed-effects variables (with statistical adjustments for the known violations of the linear regression assumptions). Changing the size of the emission zone did not seem to have a strong effect on the model fit (as determined by MSE); the inclusion of interaction terms representing interactions between

temperature and emissions resulted in more noticeable changes in the model fit, particularly in the southern regions of the study area. This made some physical sense, because the progression of chemical reactions that produces $PM_{2.5}$ from SO_2 emissions is known to be related to temperature and humidity.

Of the 42 models shown in Table 11 of the Investigators' Report, the model that the team selected for further evaluation contained the main SO_2 emission variable, an interaction variable for temperature and emissions, circular emission zones up to 400 miles from the $PM_{2.5}$ monitor, a set of dummy variables for the effect of calendar month, and the statistical adjustments for known violations of linear regression assumptions.

MODEL EVALUATION

In addition to the use of MSE as a measure of model fit, Morgenstern and colleagues also made some external comparisons of the models built for the study. As an initial check, they rebuilt the model using data for the year 2006, which was not included in the data used to build their preferred model. They compared results for the 609 $PM_{2.5}$ monitors that met the data quality requirements described above, as well as for two subsets of these monitors (i.e., 164 monitors that were also in the 1999–2005 dataset and the remaining 445 monitors). The results, presented in Table 13 of the Investigators' Report, showed substantially lower MSE values (29% reduction in MSE for the monitors in the original dataset; 24.5% for the others) for both subsets of the 2006 dataset than those for the 1999–2005 dataset. This indicates that the investigators' source–receptor model more precisely predicted the monthly values of $PM_{2.5}$ measured at the monitors, given the emissions data and fixed-effects variables, for the 2006 data than for the 1999–2005 data.

As a further test, the team compared the performance of the source–receptor model with the results of the EPA's Regulatory Impact Analysis for the Clean Air Interstate Rule (CAIR), which used the CMAQ model to predict changes in ambient $PM_{2.5}$. Using this model, the EPA predicted that the expected reductions in SO_2 from CAIR would result in reductions in $PM_{2.5}$ averaged across the eastern United States of 0.73 $\mu g/m^3$ by 2010 and 0.89 $\mu g/m^3$ by 2015. On a population-weighted basis, this translated to a reduction of 0.266 $\mu g/m^3$ per ton of SO_2 by 2010 and 0.290 $\mu g/m^3$ per ton of SO_2 by 2015. Using their preferred model, Morgenstern and his team estimated that reductions in $PM_{2.5}$ under Title IV of the Clean Air Act were equivalent to a reduction of 0.380 $\mu g/m^3$ per ton of SO_2 by 2005. Although these estimates are difficult to compare directly because of their different models of origin, regulatory environments, datasets, and time periods, the predictions were not extremely different.

COUNTERFACTUAL ANALYSIS

In order to evaluate the impact of the emissions reductions that occurred under Title IV of the Clean Air Act, the investigators also evaluated a counterfactual scenario to simulate a situation in which no mandated reductions in SO_2 had occurred over the time period of the study. This evaluation assumed that electric power plants covered under the Acid Rain Program continued to emit SO_2 at the same rate per million BTUs that they did prior to the implementation of the regulations and that electric power consumption would have been the same as the actual consumption for the time period. They then ran the same preferred model, built from the source–receptor analysis of actual 1999–2000 SO_2 emissions and ambient $PM_{2.5}$ data, with the counterfactual SO_2 emissions data and compared the counterfactual results with the actual measurements.

KEY FINDINGS

Morgenstern and colleagues succeeded in fitting a source–receptor model using the available SO_2 emissions compliance data and the Air Quality System $PM_{2.5}$ monitoring data for 1999–2005. Their preferred model, which included a temperature–emissions interaction term and monthly dummy variables, predicted monthly average $PM_{2.5}$ concentrations at the Air Quality System monitoring locations with reasonable precision (MSE = 9.37 for emission zones within a 400-mile radius, indicating a mean prediction error at a "typical" monitor of $\sqrt{9.37} = \pm 3.06$ $\mu g/m^3$).

The preferred model performed even more favorably with the data from 2006, which included 445 more monitoring stations than the 164 in the 1999–2005 group. When applied to 2006 data from the 164 monitors, the model produced more precise predictions (MSE = 7.85 for 2006 versus 9.37 for the 1999–2005 data). When the model was applied to 2006 data for the remaining 445 monitors that did not provide data for the original model-building exercise, precision improved even further (MSE = 7.19).

Because a primary goal of the study was to estimate the impact of the EPA's regulatory efforts under Title IV of the Clean Air Act, Morgenstern and colleagues constructed a counterfactual scenario in which there was no additional regulation of SO_2 and NO_x emissions from power plants, power consumption followed the same patterns that it did in reality, and the amount of SO_2 released per million BTUs for each power plant remained at pre-regulatory levels. Comparing the counterfactual scenario with the actual

regulatory scenario, the investigators calculated that Title IV of the Clean Air Act resulted in an estimated reduction in ambient $PM_{2.5}$ concentrations (averaged across the eastern United States) of 1.07 μg/m^3 between 1999 and 2005 (or 0.89 μg/m^3 on a population-weighted basis).

HEI REVIEW COMMITTEE EVALUATION OF THE REPORT

The HEI Health Review Committee independently reviewed the report by Morgenstern and colleagues and found it to be a thoughtful and detailed effort to apply data-driven models to answering the difficult question of whether a specific intervention resulted in reduced levels of ambient air pollution. The Committee's review focused on several technical areas of the project that were important to assessing the validity and utility of the effort, including the data used to build the models, the methods and statistical aspects of the model building, and an evaluation of the selected models. The Committee also assessed the investigators' modeling work and findings for their scientific impact on the air quality modeling field and impact with respect to the objectives of HEI's Accountability Program, under which this project was funded. Specific issues raised during the Committee's critical review are detailed below.

DATA ISSUES

The Committee found the brief description of the datasets and their construction, with accompanying summary statistics, in the Investigators' Report to be helpful for understanding the emissions characteristics and receptor data. However, it noted that the investigators performed both adjustments and imputations to these datasets in order to "balance" (i.e., fill holes and gaps in) the data. For example, data were imputed for 25 of the 193 monitor sites in the main dataset for an unknown number of observations. The Committee felt that the report lacked a stepwise description of how each addition or change in the datasets, such as modifying the emissions data from the CAM units with information from the NEI database, might have affected the modeling results. This expansion of the base dataset without systematic evaluation of the consequences might have affected the study's results and scientific conclusions and thus reduces their interpretability.

STUDY DESIGN ISSUES

The Committee outlined a number of issues relating to the methods used by the investigators in their modeling efforts. Although the investigators intended to use "relatively transparent statistical methods," their descriptions of data handling, statistical approach, and modeling were in many places difficult to follow. The Committee found that the statistical tools used in the report were sound for assessing associations but that the complicated implementation made it difficult to infer causation. Even though the investigators said they intended to develop a new method to assess the impact of changes in emissions on ambient concentrations using a statistical approach, the Committee felt that the investigators could have done more to explain the choices they made and the potential impacts of those choices on the resulting models. The Committee's more extensive concerns with specific aspects of the investigators' methods are discussed in this section.

Excess Model Parameters

Statistical modeling values parsimony — the use of a minimal number of explanatory variables — for a variety of technical reasons. Each additional variable added to a model, although potentially useful to control confounding or improve prediction, can also introduce interaction problems with other variables (such as correlations) that can destabilize the model. Each additional variable, if it is not necessary to improve model fit or control confounding, can also reduce the ability of the model to predict the outcome variable properly should the model be applied to a new dataset. Ideally, variables should not be added in large numbers based on a priori assumptions without some evaluation of how each variable performs in the model.

The inclusion of a large number of variables in a model is particularly problematic when the number is high compared with the number of observations in the dataset. In the current study, some models attempted to estimate more than 1300 parameters for approximately 16,000 observations. The investigators' attempts to control for unmeasured confounding by including large numbers of variables and interaction terms thus came at a cost to the interpretability and generalizability of the model — problems to be avoided when developing a model for wider applications. Although Morgenstern has indicated to the Committee that the inclusion of a large number of variables to represent unknown factors out of concern for omitted-variable bias is a standard practice in the type of econometric approach that he and his team applied here, the Committee's concerns about the statistical consequences remain.

The Committee noted that even with the large number of parameters, the results from the models exhibited residual spatial dependence, indicating the need for a more carefully designed and tested scheme for selecting the parameters used to build the model. Although the investigators attempted to adjust for this spatial dependence by

incorporating models for spatial and temporal dependence, it was not clear to the Committee how the dependencies were accounted for in the model assessments.

Model Evaluation

A primary goal of this study was to create a model of air quality changes subsequent to the implementation of regulations to reduce emissions that could be adapted to data-driven air quality modeling in similar situations. The Committee was thus concerned about the extent of model evaluation efforts and the types of evaluations performed by Morgenstern and colleagues.

The investigators did not perform a "leave one out" evaluation of their models, a common method of model evaluation where observations are omitted one by one in turn, estimated using the remaining data in the model, and then compared with the estimates to assess the predictive value of the model. The Committee also noted that the findings of the study would have been strengthened by a more detailed evaluation of the new models using out-of-sample predictions. It was unclear how the investigators applied their 1999–2005 models to 2006 data for the out-of-sample predictions they reported, given that their models had multiple fixed-effects terms for year and site that would preclude a direct application to data from other years and sites. Although the Committee found it helpful that the investigators did provide additional information in their report to clarify their comparisons between the 1999–2005 and 2006 models, it would have been useful had the investigators been able to evaluate their data-driven models side by side with the CMAQ models for the same years and areas and to directly compare the results and evaluation statistics. During its review of the report, the Committee commented that the investigators should concentrate on what their approach provides that the CMAQ model does not. The Committee noted that a key advantage of the approach presented in the report is that it provides a receptor-oriented response to emission changes, which the CMAQ model does not currently do for most PM species, and that this receptor approach is based on actual pollution measurements. Although one might argue that having a statistically based, data-driven approach is advantageous because it can be applied in areas where a CMAQ model simulation is more challenging (e.g., in less-developed parts of the world), this statistical approach is very data-intensive and hence would suffer challenges of a magnitude similar to those of chemical transport models.

RESULTS

The Committee's principal concerns with the results of the study centered on their interpretability, given the issues noted above with the building of the models and their evaluation. The inclusion of a high number of parameters compared with the number of observations and the calculation of error based on another model (the CMAQ model) led the Committee to express uncertainty about the estimates resulting from the model. Although omitted-variable bias might be a problem, the number of observations compared with the number of variables is still mathematically problematic and reduces confidence in the model results.

Although the Committee noted that the investigators made some observations about their results with regard to their predictions of pollutant concentrations that are consistent with the real-world behavior of pollutants, their assessments were somewhat speculative. For example, the investigators noted the spatial patterns of predicted concentrations of pollutants as compared with the distance from the sources and magnitude of emissions as an indication that their models were performing in an expected manner, but they did not cite the available air quality monitoring literature on field measurements surrounding coal-burning power plants.

Certainty of Results

The Committee noted that one key finding of the study was that reduced emissions from power plants as a result of the implementation of regulations to curb acid rain were also associated with lower PM in areas immediately surrounding the power plants (out to a 400-mile radius). The Committee, however, remains concerned about the interpretability of the investigators' numerical estimates from these models. During the review process, the Committee pointed out that inference depends on understanding the roles of the terms in the models, which is difficult when the models rely on a large number of dummy variables. The Committee believed that it would have been useful to understand how the estimates differed according to various factors (e.g., season and calendar year) and what additional insights this information might yield about the effects of emissions reductions on estimated PM concentrations in the area. Similarly, the Committee thought that the discussion of model uncertainty was of far greater importance than the discussion of parameter uncertainty, because the study was attempting to develop a new model for wider application.

Finally, the Committee was concerned that possible residual confounding remained in either or both of the estimates detailed in the report. Despite the large — perhaps excessive — number of variables included in each model, the Committee did not see that there was sufficient evaluation of the issues of potential confounding in the analyses as presented.

CONCLUSIONS

Overall, the HEI Review Committee found that the primary strength of the study was that it was data-driven and observational, rather than simply using complex modeling techniques to assess the impact of a regulatory intervention. Model development was focused on building models of air quality that reflected pollutant concentrations at a distance from an emitting facility and that were based on measured industrial outputs and measured concentrations of pollutants. The investigators' approach was derived from a statistical approach that was somewhat outside the usual disciplines encountered in air pollution research, bringing a different perspective to the work. The estimates obtained through these models seemed to be within reason in terms of their magnitude and level of predictive error, and the Committee believed that the authors were appropriately careful about the inferences they drew from their work and reasonably cautious about their findings. It also found that Morgenstern and his team made a good-faith effort to address the general scientific questions about accountability (i.e., the effects of regulations on air quality) that they originally intended to address.

The Committee expressed a number of concerns about the development of the models and their potential application to datasets other than those used to build them. It found that the modeling process, as described, was difficult to follow and that it was somewhat difficult to understand the investigators' model evaluation efforts and to interpret the numerical results of the models. Because the forms of the final models were driven by the nature of the data, the Committee noted that there was an implicit specificity of the method to the investigators' scenario and types of data, something the authors acknowledged was an important limitation of their work. Although the Committee felt there was some value to the investigators' overall approach, the concerns noted in this commentary made it difficult for the Committee to fully assess its potential application to air quality management or impact assessment.

Despite these limitations, it was the Committee's judgment that the investigators' work contributed to the discussion of what portions of PM reductions can be attributed to an emissions reduction program, with an approach that might be a useful alternative to atmospheric models in some applications. The Committee also noted that a model that estimates ambient air quality changes secondary to emissions changes might also be applied to estimating changes in criteria pollutants secondary to regulations aimed at reducing industrial emissions or even greenhouse gas emissions. The model might also be used to estimate the impacts of changing fuel sources at individual facilities (e.g., a coal-fired plant converting to natural gas) on concentrations of criteria pollutants in the area. Furthermore, this research and these models might be able to provide useful information to some organizations that want a quick estimate of how much specific emitting facilities affect specific pollutant monitors or even communities, although atmospheric pollution dispersion models such as CALPUFF might be more readily applied than the approach developed here.

ACKNOWLEDGMENTS

The Health Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigators' Report. The Committee is also grateful to Aaron Cohen for his oversight of the study, to Kate Adams for her assistance in preparing its Commentary, to George Simonson for his science editing of the Investigators' Report and its Commentary, and to Suzanne Gabriel, Barbara Gale, Fred Howe, Bernard Jacobson, and Flannery Carey McDermott for their roles in preparing this Research Report for publication.

REFERENCES

Health Effects Institute. 2000a. Special Report: Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Health Effects Institute, Cambridge, MA.

Health Effects Institute. 2000b. HEI Strategic Plan for Understanding the Health Effects of Air Pollution 2000–2005. Health Effects Institute, Cambridge, MA.

Health Effects Institute. 2009. Proceedings of an HEI Workshop on Further Research to Assess the Health Impacts of Actions Taken to Improve Air Quality. Communication 13. Health Effects Institute, Boston, MA.

HEI Accountability Working Group. 2003. Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research. Communication 11. Health Effects Institute, Boston, MA.

Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA III, Thurston G, Calle EE, Thun MJ. 2009. Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality. Research Report 140. Health Effects Institute, Boston, MA.

Lepeule J, Laden F, Dockery D, Schwartz J. 2012. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard Six Cities study from 1974 to

2009. Environ Health Perspect 2012 Mar 28 [Epub ahead of print].

Lin C. 2010. A spatial econometric approach to measuring air pollution externalities. J Reg Anal Pol 40(1):1–19.

Pope CA III, Ezzati M, Dockery DW. 2009. Fine-particulate air pollution and life expectancy in the United States. N Engl J Med 360:376–386.

U.S. Environmental Protection Agency. 1990. The Clean Air Act Amendments of 1990: Overview. EPA, Washington, DC. Available from *www.epa.gov/air/caa/overview.txt.* Accessed 5/31/2012.

U.S. Environmental Protection Agency. 1998. Particulate Matter ($PM_{2.5}$) Speciation Guidance Document. U.S. Environmental Protection Agency Monitoring and Quality Assurance Group, Emissions, Monitoring, and Analysis Division, Office of Air Quality Planning and Standards, Research Triangle Park, NC. Available from *www.epa.gov/ttn/amtic/files/ambient/pm25/spec/specpln2.pdf*. Accessed 8/23/2012.

U.S. Environmental Protection Agency. 2002. Overview of the Human Health and Environmental Effects of Power Generation: Focus on Sulfur Dioxide (SO_2), Nitrogen Oxides (NO_x) and Mercury (Hg). EPA, Washington, DC. Available from *www.epa.gov/clearskies/pdfs/overview.pdf*. Accessed 5/31/2012.

U.S. Environmental Protection Agency. 2012. Community Multiscale Air Quality. EPA, Research Triangle Park, NC. Available from *www.epa.gov/asmdnerl/CMAQ/index.html*. Accessed 8/20/2012.

RELATED HEI PUBLICATIONS: ASSESSING EXPOSURE TO AIR POLLUTION

Number	Title	Principal Investigator	Date
Research Reports			
167	Assessment and Statistical Modeling of the Relationship Between Satellite-Based Estimates and Measurements of $PM_{2.5}$ in the Eastern United States	C.J. Paciorek	2012
161	Assessment of the Health Impacts of Particulate Matter Characteristics	M.L. Bell	2012
158	Air Toxics Exposure from Vehicle Emissions at a U.S. Border Crossing: Buffalo Peace Bridge Study	J.D. Spengler	2011
152	Evaluating Heterogeneity in Indoor and Outdoor Air Pollution Using Land-Use Regression and Constrained Factor Analysis	J.I. Levy	2010
142	(APHENA) Air Pollution and Health: A European and North American Approach	K. Katsouyanni J.M. Samet	2009
140	Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality	D. Krewski	2009
139	Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: The NLCS-AIR Study	B. Brunekreef	2009
HEI Communications			
14	HEI's Research Program on the Impact of Actions to Improve Air Quality: Interim Evaluation and Future Directions		2009
11	Assessing the Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research		2003
10	Improving Estimates of Diesel and Other Emissions for Epidemiologic Studies		2003
7	Diesel Workshop: Building a Research Strategy to Improve Risk Assessment		1999
Special Reports			
17	Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects	HEI Panel on the Health Effects of Traffic-Related Air Pollution	2010
	Particulate Air Pollution and Daily Mortality: Analyses of the Effects of Weather and Multiple Air Pollutants: The Phase 1.B Report of the Particulate Epidemiology Evaluation Project	J.M. Samet	1997

Copies of these reports can be obtained from the Health Effects Institute and many are available at http://pubs.healtheffects.org.

HEI BOARD, COMMITTEES, and STAFF

Board of Directors

Richard F. Celeste, Chair *President Emeritus, Colorado College*

Sherwood Boehlert *Of Counsel, Accord Group; Former Chair, U.S. House of Representatives Science Committee*

Enriqueta Bond *President Emerita, Burroughs Wellcome Fund*

Purnell W. Choppin *President Emeritus, Howard Hughes Medical Institute*

Michael T. Clegg *Professor of Biological Sciences, University of California–Irvine*

Jared L. Cohon *President, Carnegie Mellon University*

Stephen Corman *President, Corman Enterprises*

Gowher Rizvi *Vice Provost of International Programs, University of Virginia*

Linda Rosenstock *Dean Emerita and Professor of Health Policy and Management, Environmental Health Sciences and Medicine, University of California–Los Angeles*

Henry Schacht *Managing Director, Warburg Pincus; Former Chairman and Chief Executive Officer, Lucent Technologies*

Warren M. Washington *Senior Scientist, National Center for Atmospheric Research; Former Chair, National Science Board*

Archibald Cox, Founding Chair *1980–2001*

Donald Kennedy, Vice Chair Emeritus *Editor-in-Chief Emeritus,* Science*; President Emeritus and Bing Professor of Biological Sciences, Stanford University*

Health Research Committee

David L. Eaton, Chair *Associate Vice Provost for Research and Director, Center for Ecogenetics and Environmental Health, School of Public Health, University of Washington–Seattle*

David Christiani *Elkan Blout Professor of Environmental Genetics, Harvard School of Public Health*

David E. Foster *Phil and Jean Myers Professor, Department of Mechanical Engineering, Engine Research Center, University of Wisconsin–Madison*

Uwe Heinrich *Professor, Medical School Hannover; Executive Director, Fraunhofer Institute for Toxicology and Experimental Medicine, Hanover, Germany*

Grace LeMasters *Professor of Epidemiology and Environmental Health, University of Cincinnati College of Medicine*

Sylvia Richardson *Professor and Director, MRC Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom*

Allen L. Robinson *Professor, Departments of Atmospheric Science and Mechanical Engineering, Colorado State University*

Richard L. Smith *Director, Statistical and Applied Mathematical Sciences Institute, University of North Carolina–Chapel Hill*

James A. Swenberg *Kenan Distinguished Professor of Environmental Sciences, Department of Environmental Sciences and Engineering, University of North Carolina–Chapel Hill*

HEI BOARD, COMMITTEES, and STAFF

Health Review Committee

Homer A. Boushey, Chair *Professor of Medicine, Department of Medicine, University of California–San Francisco*

Ben Armstrong *Reader in Epidemiological Statistics, Public and Environmental Health Research Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, United Kingdom*

Michael Brauer *Professor, School of Environmental Health, University of British Columbia, Canada*

Bert Brunekreef *Professor of Environmental Epidemiology, Institute of Risk Assessment Sciences, University of Utrecht, the Netherlands*

Mark W. Frampton *Professor of Medicine and Environmental Medicine, University of Rochester Medical Center*

Stephanie London *Senior Investigator, Epidemiology Branch, National Institute of Environmental Health Sciences*

Armistead Russell *Georgia Power Distinguished Professor of Environmental Engineering, School of Civil and Environmental Engineering, Georgia Institute of Technology*

Lianne Sheppard *Professor of Biostatistics, School of Public Health, University of Washington–Seattle*

Officers and Staff

Daniel S. Greenbaum *President*
Robert M. O'Keefe *Vice President*
Rashid Shaikh *Director of Science*
Barbara Gale *Director of Publications*
Jacqueline C. Rutledge *Director of Finance and Administration*
Helen I. Dooley *Corporate Secretary*

Kate Adams *Senior Scientist*
Aaron J. Cohen *Principal Scientist*
Maria G. Costantini *Principal Scientist*
Philip J. DeMarco *Compliance Manager*
Suzanne Gabriel *Editorial Assistant*
Hope Green *Editorial Assistant (part time)*
L. Virgi Hepner *Senior Science Editor*
Anny Luu *Administrative Assistant*
Francine Marmenout *Senior Executive Assistant*
Nicholas Moustakas *Policy Associate*
Hilary Selby Polk *Senior Science Editor*
Jacqueline Presedo *Research Assistant*
Sarah Rakow *Science Administrative Assistant*
Evan Rosenberg *Staff Accountant*
Robert A. Shavers *Operations Manager*
Geoffrey H. Sunshine *Senior Scientist*
Annemoon M.M. van Erp *Managing Scientist*
Katherine Walker *Senior Scientist*

HEALTH
EFFECTS
INSTITUTE

101 Federal Street, Suite 500
Boston, MA 02110, USA
+1-617-488-2300
www.healtheffects.org

RESEARCH
REPORT

Number 168
November 2012