



## RESEARCH REPORT

HEALTH  
EFFECTS  
INSTITUTE

Number 187  
May 2016

### **Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations**

Corwin Matthew Zigler, Chanmin Kim, Christine Choirat,  
John Barrett Hansen, Yun Wang, Lauren Hund, Jonathan Samet,  
Gary King, and Francesca Dominici





# Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations

Corwin Matthew Zigler, Chanmin Kim, Christine Choirat,  
John Barrett Hansen, Yun Wang, Lauren Hund, Jonathan Samet,  
Gary King, and Francesca Dominici

with a Critique by the HEI Health Review Committee



Research Report 187

Health Effects Institute

Boston, Massachusetts

*Trusted Science • Cleaner Air • Better Health*

Publishing history: This document was posted at [www.healtheffects.org](http://www.healtheffects.org) in May 2016.

Citation for document:

Zigler CM, Kim C, Choirat C, Hansen JB, Wang Y, Hund L, et al. 2016. Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. Research Report 187. Boston, MA:Health Effects Institute.

© 2016 Health Effects Institute, Boston, Mass., U.S.A. Cameographics, Belfast, Me., 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

About HEI	vii
About This Report	ix
Preface	xi
HEI STATEMENT	1
INVESTIGATORS' REPORT <i>by Zigler et al.</i>	5
ABSTRACT	5
Introduction	5
Methods	5
Results	6
Conclusion	6
INTRODUCTION	6
Overview	7
Case Study 1: PM <sub>10</sub> Nonattainment Designations	7
Case Study 2: Scrubber Installations on Coal-Fired Power Plants	7
A Note on the Word "Causal"	8
SPECIFIC AIMS	8
METHODS AND STUDY DESIGN	8
Publicly Available Data and Reproducible Research for Accountability Assessment	8
Ambient Monitoring Data	9
Medicare Health Outcomes Data	9
Regulatory Data Sources	9
Other Supporting Data Sources	10
Linking Existing Data	10
The Harvard Dataverse	10
AREPA R Package	10
Statistical Perspectives for Causal Accountability Assessment	11
Direct versus Indirect Accountability	12
Potential-Outcomes Methods: Framing Observational Studies as Hypothetical (Approximate) Randomized Experiments	12
Methods for Confounding Adjustment: Propensity Scores	13
Causal Pathways Analyses: Causal Mediation Analysis and Principal Stratification	14
RESULTS	15
Case Study 1: Accountability Assessment of PM <sub>10</sub> Nonattainment Designations in the Western United States	15
Linked Data Sources	16
Defining the Intervention for Direct-Accountability Assessment: Initial PM <sub>10</sub> Nonattainment Designations	19
Potential-Outcomes Approach and Causal Effects of Interest	19
Estimation of Propensity Scores for Confounding Adjustment	20

# Research Report 187

Checking Covariate Balance	21
Models for Estimating Causal Effects	22
<b>Case Study 1: Results</b>	23
Unadjusted Comparisons	24
Average Causal Effects on Average Annual Ambient PM <sub>10</sub> in 1999–2001	24
Average Causal Effects on Medicare Health Outcomes	24
Associative and Dissociative Effects	24
<b>Conclusion and Discussion of Case Study 1</b>	26
Additional Work in Progress	29
<b>Case Study 2: Accountability Assessment of Power-Plant Emissions Controls</b>	29
Linked Data Sources	31
Defining the Intervention: SO <sub>2</sub> Scrubber Installation	32
Defining Potential Outcomes for Principal Stratification and Causal Mediation Analysis	33
Estimation: New Methods for Bayesian Nonparametric Mediation Analysis	35
<b>Case Study 2: Results</b>	35
<b>Conclusion and Discussion of Case Study 2</b>	41
Extension of Case Study 2 to Health-Outcomes Analyses	42
Additional Work in Progress	43
<b>DISCUSSION AND CONCLUSIONS</b>	43
<b>ACKNOWLEDGMENTS</b>	44
<b>REFERENCES</b>	44
<b>HEI QUALITY ASSURANCE STATEMENT</b>	47
<b>MATERIALS AVAILABLE ON THE WEB</b>	48
<b>ABOUT THE AUTHORS</b>	48
<b>OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH</b>	49
<b>ABBREVIATIONS AND OTHER TERMS</b>	49
<b>CRITIQUE</b> <i>by the Health Review Committee</i>	51
<b>INTRODUCTION AND SCIENTIFIC BACKGROUND</b>	51
<b>SUMMARY OF THE STUDY</b>	52
<b>Specific Aims</b>	52
<b>Methods</b>	53
Overview of the Causal Methods Development	53
<b>Case Study 1: Accountability Assessment of PM<sub>10</sub>     Nonattainment Designations in the Western United States</b>	53
Introduction	53
Statistical Methods	54
Results	55
<b>Case Study 2: Accountability Assessment of Power-Plant     Emissions Controls</b>	55
Introduction	55
Statistical Methods	56
Results	56

## Research Report 187

---

HEALTH REVIEW COMMITTEE'S CRITIQUE	57
Case Study 1: Impacts of Designation as Nonattainment Area on PM <sub>10</sub> and Health	58
Case Study 2: Scrubber Installation on Coal-Fired Power Plants	59
SUMMARY AND CONCLUSIONS	61
ACKNOWLEDGMENTS	62
REFERENCES	62
 Related HEI Publications	 65
 HEI Board, Committees, and Staff	 67



# 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 balanced funding from the U.S. Environmental Protection Agency and 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 330 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 more than 260 comprehensive reports published by HEI, as well as in more than 1000 articles in the peer-reviewed literature.

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](http://www.healtheffects.org)), printed reports, newsletters and other publications, annual conferences, and presentations to legislative bodies and public agencies.



# ABOUT THIS REPORT

Research Report 187, *Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations*, presents a research project funded by the Health Effects Institute and conducted by Dr. Corwin M. Zigler of the Harvard T.H. Chan School of Public Health, Boston, Massachusetts, and his colleagues. The 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 Zigler and colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

**The Critique**, prepared by members of the Health Review Committee with the assistance of HEI staff, 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 Critique reflects the information provided in the final version of the report.



# PREFACE

---

## HEI's Accountability 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 *accountability research*, *outcomes 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 accountability research, *Communication 11, Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research* (HEI Accountability Working Group 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 accountability 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 Accountability Research program, which is discussed below.

Between 2002 and 2004, HEI issued four requests for applications (RFAs), under which eight studies were funded (see Table). A ninth study was funded later, under Request for Preliminary Applications (RFP) 05-3, "Health Effects of Air Pollution." Following this first wave of research, HEI held further workshops to discuss lessons learned, identify key remaining questions, and plan a second wave of research. These efforts led to the publication of *Communication 14* (van Erp and Cohen 2009) and *Communication 15* (HEI 2010b), and the issuance of RFA 11-1, "Health Outcomes Research — Assessing the Health Outcomes of Air Quality Actions." The first wave of research primarily consisted of studies evaluating relatively short-term, local-scale, and sometimes temporary interventions; RFA 11-1 solicited additional studies with a focus on longer-term, regional- and national-scale regulations, including programs targeted at improving air quality surrounding major ports, as well as further methods development.

This preface describes both the framework of accountability research as it relates to air quality regulations and HEI's Accountability 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

## Preface

---

### HEI's Accountability Research Program

---

RFA / Investigator (Institution)	Intervention	Study or Report Title
<b>First-Wave Studies<sup>a</sup></b>		
<b>RFA 02-1</b>		
Douglas Dockery (Harvard School of Public Health, Boston, MA)	Coal ban in Irish cities	Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland (Research Report 176; 2013)
Annette Peters (Helmholtz Zentrum München—German Research Center for Environment and Health, Neuherberg, Germany)	Switch from brown coal to natural gas for home heating and power plants, changes in motor vehicle fleet after reunification of Germany	The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany (Research Report 137; 2009)
<b>RFA 04-1</b>		
Frank Kelly (King's College London, U.K.)	Measures to reduce traffic congestion in the inner city of London	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 (Research Report 155; 2011)
<b>RFA 04-4</b>		
Frank Kelly (King's College London, U.K.)	Measures to exclude most polluting vehicles from entering greater London	The London Low Emission Zone Baseline Study (Research Report 163; 2011)
Richard Morgenstern (Resources for the Future, Washington, DC)	Measures to reduce sulfur emissions from power plants east of the Mississippi River	Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments (Research Report 168; 2012)
Curtis Noonan (University of Montana, Missoula, MT)	Wood stove change-out program	Assessing the Impact of a Wood Stove Replacement Program on Air Quality and Children's Health (Research Report 162; 2011)
Jennifer Peel (Colorado State University, Fort Collins, CO)	Measures to reduce traffic congestion during the Atlanta Olympics	Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes (Research Report 148; 2010)
Chit-Ming Wong (University of Hong Kong)	Measures to reduce sulfur content in fuel for motor vehicles and power plants	Impact of the 1990 Hong Kong Legislation for Restriction on Sulfur Content in Fuel (Research Report 170; 2012)
<b>RFA 05-3</b>		
Junfeng (Jim) Zhang (University of Medicine and Dentistry of New Jersey, Piscataway, NJ)	Measures to improve air quality during the Beijing Olympics	Cardiorespiratory Biomarker Responses in Healthy Young Adults to Drastic Air Quality Changes Surrounding the 2008 Beijing Olympics (Research Report 174; 2013)

*Table continues next page*

---

## Preface

### HEI's Accountability Research Program *(continued)*

RFA / Investigator (Institution)	Intervention	Study or Report Title
<b>Second-Wave Studies<sup>a</sup></b>		
<b>RFA 11-1</b>		
Frank Gilliland (University of Southern California)	California and federal programs to improve air quality, including control of emissions from diesel engines and other sources targeted at freight transport and ports, as well as stationary sources	The Effects of Policy-Driven Air Quality Improvements on Children's Respiratory Health (Report in review)
Ying-Ying Meng (University of California–Los Angeles)	2006 California Emissions Reduction Plan for Ports and Goods Movement to control emissions from road, rail, and marine transportation, focusing on the ports of Los Angeles and Long Beach	Improvements in Air Quality and Health Outcomes Among California Medicaid Enrollees Due To Goods Movements (Study ongoing)
Armistead Russell (Georgia Institute of Technology)	Programs to control emissions from major stationary sources and mobile sources in the Southeast United States	Impacts of Emission Changes on Air Quality and Acute Health Effects in the Southeast, 1993–2012 (Study ongoing)
Corwin Zigler (Harvard T.H. Chan School of Public Health)	National regulations to improve air quality focusing on State Implementation Plans for particulate matter	Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations (Current report)

<sup>a</sup> Abbreviations: RFA, Request for Applications; RFPA, Request for Preliminary Applications

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 accountability 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; 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 regula-

tions 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 imple-

mented 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.

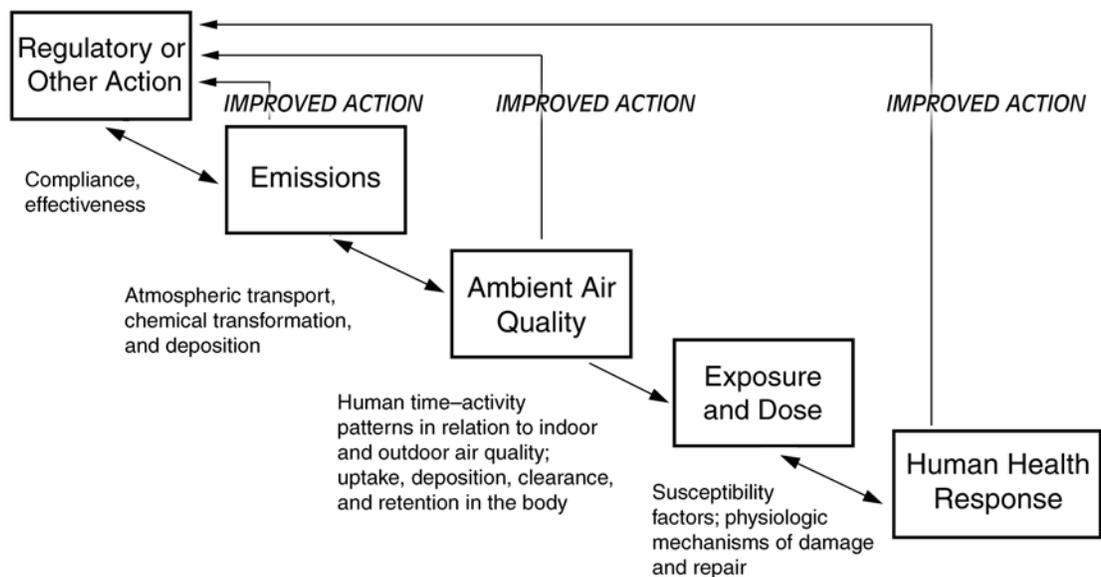
These inherent challenges are well documented in Communication 11 (HEI Accountability Working Group 2003), which was intended to advance the concept of accountability research and to foster the development of methods and studies throughout the relevant scientific and policy communities. In addition, recent advances in data collection and analytic techniques provide an unprecedented opportunity to im-

prove our assessments of the effects of air quality interventions.

### THE ACCOUNTABILITY 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 "accountability evaluation cycle" (see Figure), each stage of which affords 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 unin-



**Accountability 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.

tended 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 accountability 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 accountability 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 ACCOUNTABILITY RESEARCH PROGRAM

The first wave of HEI's Accountability Research program included nine studies (see Table). These studies involved the measurement of indicators along the entire accountability evaluation cycle, from regulatory or other interventions to human health outcomes. Many of the studies focused on interventions that were 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 also supported 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 study by Morgenstern and colleagues (2012) 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.

Subsequently, HEI funded four studies as part of the second wave of HEI's Accountability program (see Table). Two studies evaluated regulatory and other actions at the national or regional level implemented over multiple years; a third study is evaluating 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 has developed methods to support such accountability research. Gilliland and colleagues evaluated respiratory symptoms and lung function growth in children in Southern California from 1993–2012. They used data from three cohorts of the Children's Health Study, attempting to relate changes in health outcomes to major air quality regulations during that time period (Berhane et al. 2016; Gauderman et al. 2015; Lurmann et al. 2015). Russell and colleagues are assessing the effect of major stationary source and mobile source control programs on emissions and air quality in the Southeast United States, using detailed emissions and air pollution measurements and modeling combined with time-series analyses of cardiovascular and respiratory emergency department visits and hospital admissions. Meng and colleagues are evaluating the effects on air quality and health associated with the California Air Resources Board's Emission Reduction Plan for Ports and Goods Movement. They are examining the changes in criteria and hazardous air pollutants and characterizing health outcomes among Medicaid beneficiaries in the

region surrounding the ports of Long Beach and Los Angeles. Zigler and colleagues, as described in this Investigators' Report, developed and applied statistical methods to evaluate long-term regulatory actions, focusing on the Clean Air Act and the role of attainment status of counties for PM<sub>10</sub>, O<sub>3</sub>, CO, and SO<sub>2</sub> concentrations. In particular, they focused on methods targeted on the question of whether air quality and health outcomes are causally related (Zigler and Dominici 2014).

Studies on health outcomes funded by HEI to date are summarized in the Table on page xii. The first-wave studies are described in more detail in an interim evaluation of the HEI Accountability Research program (van Erp and Cohen 2009; van Erp et al. 2012).

---

### FUTURE DIRECTIONS

---

The second stage of accountability research was largely conceived during HEI's Strategic Plan for 2010 through 2015 (HEI 2010a). During the current Strategic Plan for 2015 through 2020 (HEI 2015), HEI will continue to look closely at opportunities for unique new contributions to accountability research. We envision that future studies will again focus on large-scale, complex regulations to improve air quality and will continue to develop and implement statistical methods to tackle these complicated questions.

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](http://www.ihapss.jhsph.edu)); and
2. Data from the National Particle Component Toxicity (NPACT) initiative on concentrations of components of particulate matter with an aerodynamic diameter  $\leq 2.5 \mu\text{m}$  (PM<sub>2.5</sub>) collected at or near the 54 sites in the EPA's PM<sub>2.5</sub> 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 accountability studies. The STN Web site provides scientists an opportunity to investigate specific questions about concentrations of PM<sub>2.5</sub> components and their association with adverse health effects in regions covered by the STN network and to address questions related to accountability 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 recommendations of this workshop 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. As part of the Strategic Plan 2015–2020, HEI plans to hold another workshop to discuss the future of accountability research during 2017 or 2018.

In the interim, 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. Available: <https://hei.aer.com> [accessed 11 May 2016].

Berhane K, Chang CC, McConnell R, Gauderman WJ, Avol E, Rappaport E, et al. 2016. Association of changes in air quality with bronchitic symptoms in children in California, 1993–2012. *JAMA* 315(14):1491–1501.

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.

Gauderman WJ, Urman R, Avol E, Berhane K, McConnell R, Rappaport E, et al. 2015. Association of improved air quality with lung development in children. *N Engl J Med* 372(10):905–913.

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

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

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

Health Effects Institute. 2015. *HEI Strategic Plan for Understanding the Health Effects of Air Pollution 2015–2020*. Boston, MA:Health Effects Institute.

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. Boston, MA:Health Effects Institute.

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). Available: [www.ihapss.jhsph.edu](http://www.ihapss.jhsph.edu) [accessed 11 May 2016].

Kelly F, Anderson HR, Armstrong B, Atkinson R, Barratt B, Beevers S, et al. 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. Boston, MA:Health Effects Institute.

Kelly F, Anderson HR, Armstrong B, Atkinson R, Barratt B, Beevers S, et al. 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. Boston, MA:Health Effects Institute.

Kelly FJ, Armstrong B, Atkinson R, Anderson HR, Barratt B, Beevers S, et al. 2011c. *The London Low Emission Zone Baseline Study*. Research Report 163. Boston, MA:Health Effects Institute.

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.

Lurmann F, Avol E, Gilliland F. 2015. Emissions reduction policies and recent trends in Southern California's ambient air quality. *J Air Waste Manag Assoc* 65(3):324–335.

Matte TD, Cohen A, Dimmick F, Samet J, Samat J, Yip F, et al. 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. Boston, MA:Health Effects Institute.

National Research Council (U.S.). 1998. *Research Priorities for Airborne Particulate Matter: I. Immediate*

## Preface

- Priorities and a Long-Range Research Portfolio. Washington, DC:National Academies Press.
- National Research Council (U.S.). 2002. Estimating the Public Health Benefits of Proposed Air Pollution Regulations. Washington, DC:National Academies Press.
- 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. Boston, MA:Health Effects Institute.
- 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. Boston, MA:Health Effects Institute.
- Peters A, Breitner S, Cyrus J, Stölzel M, Pitz M, Wölke G, et al. 2009. The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany. Research Report 137. Boston, MA:Health Effects Institute.
- 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. Available: [www.cdc.gov/nceh/tracking/](http://www.cdc.gov/nceh/tracking/) [accessed 11 May 2016].
- 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. Washington, DC:Office of Air and Radiation.
- 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. Boston, MA:Health Effects Institute.
- 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, et al. 2012. Impact of the 1990 Hong Kong Legislation for Restriction on Sulfur Content in Fuel. Research Report 170. Boston, MA:Health Effects Institute.
- Zeger SL, McDermott A, Dominici F, Peng R, Samet J. 2006. Internet-Based Health and Air Pollution Surveillance System. Communication 12. Boston, MA:Health Effects Institute.
- Zhang J, Zhu T, Kipen H, Wang G, Huang W, Rich D, et al. 2013. Cardiorespiratory Biomarker Responses in Healthy Young Adults to Drastic Air Quality Changes Surrounding the 2008 Beijing Olympics. Research Report 174. Boston, MA:Health Effects Institute.
- Zigler CM, Dominici F. 2014. Point: clarifying policy evidence with potential-outcomes thinking — beyond exposure–response estimation in air pollution epidemiology. *Am J Epidemiol* 180(12):1133–1140.

# HEI STATEMENT

## Synopsis of Research Report 187

### Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations

#### INTRODUCTION

The report by Dr. Corwin Zigler and colleagues, *Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations*, is the latest in a series of reports funded as part of HEI's Accountability research program. Established 15 years ago, this program has aimed at evaluating whether regulatory and other actions taken to improve air quality have resulted in the intended improvements in air quality, exposure, and health outcomes.

Zigler and his colleagues tackled a number of important questions that have remained unanswered by previous air pollution accountability research. A major goal of the study was to use both established methods and newly developed methods that would enable a "direct" accountability assessment of air pollution interventions — that is, to assess from a statistical standpoint whether the intervention had *caused* changes in pollutant levels or health outcomes. This "direct" approach contrasts with the "indirect" accountability approach in which the future health benefits of an intervention are estimated from the intervention's projected impact on future exposures combined with the exposure–response relationships derived from retrospective epidemiological studies.

As part of demonstrating their methods, the investigators applied them in two well-developed case studies of interventions designed to have long-term impacts on health, not just the shorter term interventions that have been the focus of much previous accountability research. Longer term effects of air pollution interventions on health are important because they account for the majority of the estimated benefits from improving air quality. Another important feature of this project is the investigators' development of new methods for evaluating the impacts of interventions on multiple

#### What This Study Adds

- Zigler and colleagues have provided a well-written primer on how more systematic approaches to testing of causality (i.e., through use of causal inference frameworks and methods) could be adapted to the assessment of the effects of air pollution interventions on air quality and health.
- In a major undertaking, they successfully demonstrated the use of existing and newly developed methods in two case studies of regulatory actions: the designation of counties to be in nonattainment with the National Ambient Air Quality Standards for PM<sub>10</sub> and the installation of SO<sub>2</sub> scrubbers on power plants.
- The scrubber case study provides both newly developed methods and a rare comparison of two different but analogous statistical approaches — principal stratification and causal mediation analysis — applied to the same complex multipollutant problem.
- Their work demonstrated the critical importance of involving multidisciplinary teams with detailed technical knowledge of the interventions to ensure appropriate study design and interpretation.
- The HEI Review Committee concluded that these accountability methods are an important addition to the "toolkit" and should continue to be further explored, but cannot wholly substitute for accountability assessments that rely on evidence from other scientific methods, including more traditional epidemiology analyses.

pollutants and the pathways via which the interventions and pollutants may affect exposure and health outcomes. To provide expertise on the complexities of the air pollution interventions chosen, Dr. Zigler added to their team Mr. John Bachmann, former Associate Director for Science/Policy and New Programs for the U.S. EPA's Office of Air and Radiation. Finally, in a commitment to transparency and data access, they plan to make publicly available the statistical code necessary both to assemble and link their data sources and to implement their newly developed methods.

### APPROACH

As in other published work on causal methods, the first important feature of their approach was the reframing of air pollution interventions as a hypothetical randomized experiment, analogous to a randomized clinical trial in which some subjects are randomly assigned to receive "treatment" and others receive none, the "controls." Randomized studies are considered the optimal study design for determining the efficacy, or causal influence, of treatment because randomization typically results in balance of potential confounders between the treatment and control groups.

The next important feature of their approach was to apply and extend two different but conceptually analogous methods, principal stratification and causal mediation, to investigate the importance of alternative causal pathways for the interventions. The causal pathways are the pathways through which an intervention may act to cause changes in the outcome of interest. The pathway may represent the direct effect of one factor on an outcome (e.g., air pollution on health outcomes) or may involve the mediation by some intermediate step or factor.

Principal stratification involves comparison of outcomes between key strata or groupings of the data (for example, the effects on health in areas where an intervention has caused a reduction in air pollution and those where it has not). Using this general example, it defines "associative" effects as those effects on health that occur when an intervention caused a meaningful reduction in air pollution and "dissociative" effects as the effects on health outcomes that occur when the intervention did not have a causal effect on air pollution. The size of the associative effects relative to the dissociative effects provides an indication of the relative importance of the

two pathways, in this example an indication of the intermediate role of the reduction in air pollution. Causal mediation methods are also designed to evaluate the effect of mediators or intermediate steps on an outcome of interest but in a more formal way. Using our general air pollution example, causal mediation divides the effects of an intervention into two components: (1) the "natural direct" effect, defined as the direct effect of the intervention on the outcome, and (2) the "natural indirect" effect, defined as the causal effect mediated by changes in some intermediate factor like a specific air pollutant. However, unlike in principal stratification, these two effects sum to the total effect. The authors demonstrated the use of these methods in two case studies of different regulatory interventions.

In the first case study, the authors evaluated the effect on air quality and on health outcomes of designating areas of the Western United States to be in "nonattainment" with the 1987 National Ambient Air Quality Standards for PM<sub>10</sub> in the period 1990–1995. Specifically, they examined the causal effects of these designations on ambient PM<sub>10</sub> concentrations in 1999–2001 and on all-cause mortality and on hospitalizations for cardiovascular and respiratory diseases in 2001. In the framing of the analysis like a randomized controlled experiment, the areas designated as in nonattainment are considered to be assigned to "treatment" whereas attainment areas served as "controls." Because these two groups were not actually selected via a randomized process, the authors developed and used propensity scores, an aggregate measure of multiple potential confounding factors, to identify groups of nonattainment (219) and attainment areas (276) that appeared comparable. The first step was to estimate the causal effects of nonattainment designation on PM<sub>10</sub> concentrations and on Medicare health outcomes, which they did using regression techniques.

The investigators next used principal stratification to examine whether causal effects of nonattainment designation on health outcomes were more likely than not to have occurred via causal reduction in ambient PM<sub>10</sub> concentrations. For this case study, they defined "associative" effects as the effects on health when the nonattainment designation was found to cause a reduction in ambient PM<sub>10</sub> by at least 5 µg/m<sup>3</sup>, and "dissociative" effects as the effects on health outcomes that occurred when the designation did not have a causal effect on PM<sub>10</sub>.

The second case study was designed to evaluate the causal impacts on emissions and ambient  $PM_{2.5}$  of installing a range of scrubber technologies on coal-fired power plants pursuant to requirements to reduce emissions of multiple pollutants ( $SO_2$ ,  $NO_x$ , and  $CO_2$ ) under the Acid Rain Program, a program created by the 1990 amendments to the Clean Air Act. The effects of scrubbers on pollutant emissions and ambient  $PM_{2.5}$  concentrations have been well-studied and understood, so this intervention provided a good opportunity to test whether the new methods would perform as expected.

The investigators estimated the causal effect of scrubber installation on emissions by comparing the 2005 emissions of  $SO_2$ ,  $NO_x$ , and  $CO_2$  and levels of ambient  $PM_{2.5}$  observed for 63 power plants that were equipped with scrubbers (“treated”) with the emissions from those 195 power plants that were not (“controls”). Zigler and colleagues then applied both principal stratification and causal mediation methods to evaluate the extent to which the causal effect of a scrubber on ambient  $PM_{2.5}$  was mediated through reduced emissions of  $SO_2$ ,  $NO_x$ , and  $CO_2$ . This analysis is notable because it involves assessment of the roles of multiple pollutants whereas most accountability assessments consider only one. It is also a rare application of the two methods to the same complex problem.

The principal stratification analysis compared the “associative” effects of scrubbers on  $PM_{2.5}$  — the causal effects of a scrubber on ambient  $PM_{2.5}$  among power plants where emissions of  $SO_2$ ,  $NO_x$ , and  $CO_2$  were causally affected by the presence of a scrubber — with the “dissociative” effects — the causal effects of a scrubber on ambient  $PM_{2.5}$  among power plants where the emissions were not meaningfully affected.

Zigler and colleagues developed new Bayesian nonparametric methods to conduct their multipollutant causal mediation analysis. In this case study, they defined the “natural direct” effect as the effect that the presence of scrubbers had on  $PM_{2.5}$  and the “natural indirect” effects as the causal effects on  $PM_{2.5}$  mediated by changes in the emissions of the three pollutants, either individually or in various combinations with each other.

### RESULTS AND INTERPRETATION

In their evaluation of the effect of nonattainment designation, the authors concluded that there was

some evidence that the intervention caused a small reduction, on average, in ambient  $PM_{10}$  levels, in all-cause mortality, and in hospitalizations for respiratory disease among Medicare beneficiaries. They did not find a reduction in hospitalizations for cardiovascular disease.

With their principal stratification analysis, Zigler and colleagues found differing results for the intermediary role of  $PM_{10}$  in causal effects on the three health outcomes. Contrary to expectations, their analysis suggested a reduction, on average, in mortality even in areas where their analyses reported that  $PM_{10}$  was not causally affected. The authors suggested that the observed causal effect of nonattainment designation on mortality, in the absence of a strong associative effect for  $PM_{10}$ , may be due to causal pathways other than the one involving reduction of  $PM_{10}$ . However, they suggested their results provide evidence that  $PM_{10}$  played a causal role in the reduction of hospitalization for respiratory disease, but again, not for cardiovascular disease.

As the authors noted, all of the estimates from these analyses were accompanied by substantial uncertainty, indicated by broad posterior 95% confidence intervals that included zero. As a result, the HEI Health Review Committee thought the investigators generally overstated the average causal effects of nonattainment designation and the role of  $PM_{10}$  in this study. The Committee agreed that a major contributor to the uncertainty in the results was the ambiguity of the intervention; that is, that nonattainment designation is not a discrete intervention, but is subject to a number of sources of heterogeneity in the actions implemented over space and time.

In their second case study, Zigler and colleagues found results that were consistent with what is known about scrubbers. They estimated that installation of scrubbers had, on average, caused reductions in  $SO_2$ , but not in  $NO_x$  and  $CO_2$  emissions, and had also caused modest reductions in ambient  $PM_{2.5}$  concentrations. Their multipollutant causal pathways analyses using principal stratification and causal mediation methods yielded broadly similar results. That is, both led the authors to conclude that the observed causal reductions in ambient  $PM_{2.5}$  among power plants equipped with scrubbers were effected principally through the causal reduction of  $SO_2$  emissions rather than through reductions in emissions of  $NO_x$  and  $CO_2$ . Their causal mediation analysis provided a somewhat clearer

support for that conclusion because the reduction in  $PM_{2.5}$  mediated by  $SO_2$  (the natural indirect effect) was statistically significant and larger than those mediated either by  $NO_x$  and  $CO_2$ , which were all close to zero. The 95% posterior intervals for all the results in the principal stratification analysis were quite broad and included zero.

Although the scrubber case study was conceptually clearer for demonstrating the methods, the authors had made a number of simplifying assumptions that could have contributed to uncertainties in the results, a question that could be explored more fully in future analyses. The investigators' first iteration of the analysis yielded results that ran counter to established knowledge (i.e., the results suggested  $SO_2$  scrubbers' effects on ambient  $PM_{2.5}$  were not causally mediated by changes in  $SO_2$  emissions) that led them to identify and correct for additional important characteristics in their final analysis. It is still difficult to know if there were other regulation-related activities undertaken that blurred the distinctions between treated and untreated facilities and that could explain the high degree of uncertainty observed in the results.

### CONCLUSIONS

The Committee concluded that Zigler and his colleagues provided a well-conducted study and a well-written report that makes a major contribution to the field of accountability research in the context of air pollution and health. The statistical framework described in this report provides a particularly clear and explicit approach to thinking about the health impacts of all kinds of interventions designed

to reduce emissions and ambient air pollution. Although most of the causal inference methods Zigler and colleagues used were not new, their extensions to two substantive air pollution interventions and to multiple pollutants were a major undertaking in and of themselves. The advances they made in applying the methods in real applications have moved us further than other methodological studies and provided a clearer path toward further development and deployment of the methods in other settings.

What the considerable methodological work in this study indicates, however, is that the presence of a clear causal framework is not a substitute for detailed consideration of potentially important covariates and the testing of the sensitivity of results to key assumptions made in implementing the methods. Both these case studies demonstrated the critical importance of involving multidisciplinary teams with detailed technical knowledge of the interventions being studied. Even so, it is difficult to be sure to what extent the uncertainty in the causal effects estimated is attributable to weakness in the causal relationship or to the imprecision in the problem definition and underlying data. Finally, not all questions can necessarily be addressed in a causal framework, for example, situations in which suitable "controls" do not exist or in which analysts need to predict the potential impacts of some future intervention. The Committee concluded that these and other "direct" accountability methods are an important addition to the "toolkit" and should continue to be further explored, but cannot wholly substitute for "indirect" accountability methods.

## Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations

Corwin Matthew Zigler,<sup>1</sup> Chanmin Kim,<sup>1</sup> Christine Choirat,<sup>1</sup> John Barrett Hansen,<sup>2</sup> Yun Wang,<sup>1</sup> Lauren Hund,<sup>3</sup> Jonathan Samet,<sup>4</sup> Gary King,<sup>2</sup> and Francesca Dominici<sup>1</sup>

<sup>1</sup>Harvard T.H. Chan School of Public Health; <sup>2</sup>Harvard University; <sup>3</sup>Sandia National Laboratories; <sup>4</sup>University of Southern California Keck School of Medicine

---

### ABSTRACT

---

#### INTRODUCTION

The regulatory and policy environment surrounding air quality management warrants new types of epidemiological evidence. Whereas air pollution epidemiology has typically informed previous policies with estimates of exposure–response relationships between pollution and health outcomes, new types of evidence can inform current debates about the actual health impacts of air quality regulations. Directly evaluating specific regulatory strategies is distinct from and complements estimating exposure–response relationships; increased emphasis on assessing the effectiveness of well-defined regulatory interventions will enhance the evidence supporting policy decisions. The goal of this report is to provide new analytic perspectives and statistical methods for what we refer to as “direct”-accountability assessment of the effectiveness of specific air quality regulatory interventions. Toward this end, we sharpened many of the distinctions surrounding accountability assessment initially raised by the HEI Accountability Working Group (2003) through discussion, development, and deployment of statistical methods for drawing causal

inferences from observational data. The methods and analyses presented here are unified in their focus on anchoring accountability assessment to the estimation of the causal consequences of well-defined actions or interventions. These analytic perspectives are discussed in the context of two direct-accountability case studies pertaining to four different links in the so-called chain of accountability, the related series of events leading from the intervention to the expected outcomes (see Preface; HEI Accountability Working Group 2003).

#### METHODS

The statistical methods described in this report consist of both established methods for drawing causal inferences from observational data and newly developed methods for assessing causal accountability. We have sharpened the analytic distinctions between studies that directly evaluated the effectiveness of specific policies and those that estimated exposure–response relationships between pollution and health. We emphasized how a potential-outcomes paradigm for causal inference can elevate policy debates by means of more direct evidence of the extent to which complex regulatory interventions affect pollution and health outcomes. We also outlined the potential-outcomes perspective and promoted its use as a means to frame observational studies as approximate randomized experiments. Our newly developed methods for assessing causal accountability draw on propensity scores, principal stratification, causal mediation analysis, spatial hierarchical models, and Bayesian estimation.

The first case study made use of health outcomes among approximately four million Medicare beneficiaries living in the Western United States to estimate the causal health impacts of areas designated as being in nonattainment for particulate matter  $\leq 10 \mu\text{m}$  in aerodynamic diameter

---

This Investigators' Report is one part of Health Effects Institute Research Report 187, which also includes a Critique by the Health Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. Corwin Matthew Zigler, Harvard T.H. Chan School of Public Health, Building 2, 4th Floor, 655 Huntington Ave., Boston, MA 02115; e-mail: [zigler@hsph.harvard.edu](mailto:zigler@hsph.harvard.edu).

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR-83467701 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.

(PM<sub>10</sub>\*) according to the 1987 National Ambient Air Quality Standards (NAAQS). The second case study focused on developing and testing our new, advanced methodology for multipollutant accountability assessment by examining the extent to which sulfur dioxide (SO<sub>2</sub>) scrubbers on coal-fired power plants causally affect emissions of SO<sub>2</sub>, nitrogen oxides (NO<sub>x</sub>), and carbon dioxide (CO<sub>2</sub>) as well as the extent to which emissions reductions mediate the causal effect of a scrubber on ambient concentrations of PM<sub>2.5</sub>.

Both case studies were anchored in our compilation of national, linked data on ambient air quality monitoring, weather, population demographics, Medicare hospitalization and mortality outcomes, continuous-emissions monitoring for electricity-generating units (EGUs) in power plants, and a variety of regulatory control interventions. The resulting database has unprecedented accuracy and granularity for conducting the types of accountability assessments presented in this report. A key component of our work was the creation of tools to help distribute our linked database and to facilitate reproducible research.

### RESULTS

In the first case study, we focused on illustrating the most fundamental features of a causal-inference perspective on direct-accountability assessment. The results indicated that all-cause Medicare mortality and respiratory-related hospitalization rates were causally reduced in areas designated as nonattainment for PM<sub>10</sub> during 1990 to 1995 compared with the rates that would have occurred without the designation.

In the second case study, which examined power-plant emissions and illustrated our newly developed statistical methods, the results indicated that the presence of an SO<sub>2</sub> scrubber causally reduced ambient PM<sub>2.5</sub> and that this reduction was mediated almost entirely through causal reductions in SO<sub>2</sub> emissions. The results were interpreted in light of the well-documented relationships between scrubbers, power-plant emissions, and PM<sub>2.5</sub>.

### CONCLUSION

By grounding accountability research in a potential-outcomes framework and applying our new methods to our collection of national data sets, we were able to provide additional sound evidence of the health effects of long-term, large-scale air quality regulations. This additional, rigorous evidence of the causal effects of well-defined actions augments the existing body of research and ensures that the

highest-level epidemiological evidence will continue to support regulatory policies. Ultimately, our research contributed to the evidence available to support to the U.S. Environmental Protection Agency (U.S. EPA) and other stakeholders for incorporating health outcomes research into policy development.

---

### INTRODUCTION

---

The claim that exposure to ambient air pollution is harmful to human health is hardly controversial in this day and age, in large part because of the evidence amassed through decades of air pollution epidemiological research. This body of research historically focused on hazard identification and more recently on estimation of exposure–response (or, more formally, concentration–response) functions relating variations in health outcomes with spatial or temporal variations in ambient pollution exposures (Chen et al. 2013; Correia et al. 2013; Dockery et al. 1993; Friedman et al. 2001; Krewski et al. 2003; Laden et al. 2006; Pope et al. 1996, 2009; Zeger et al. 2008). Although considerable uncertainty remains about essential finer-grade issues (such as the specific shape of the exposure–response functions, determinants of susceptibility, mechanics of exactly how pollution harms the human body, and the achievement of an “adequate margin of safety” as required by the U.S. Clean Air Act [CAA]), evidence of the exposure–response relationships between pollution and health has motivated a vast array of air quality control policies in the United States and abroad. The collective impact of these policies has undeniably improved ambient air quality over the past several decades (Samet 2011; U.S. EPA 2009).

Despite the success of these policies in improving air quality, an evolving regulatory, policy, and political environment is placing new demands on input from the scientific community. With the prospect of increasing costs resulting from proposed tightening of air quality standards, the evidence motivating the policies is being subjected to unprecedented scrutiny, and the scientific community must adapt by providing new and more relevant types of evidence to support current and future regulatory strategies (Dominici et al. 2014; Samet 2011; Zigler and Dominici 2014). Policymakers, legislators, industry, and the public increasingly question whether more stringent health-based standards will provide increased benefits, whether the costs are justified, and which existing strategies have provided the greatest health benefits. These considerations reflect a shift toward demanding evidence of the effectiveness of specific regulatory interventions. Starting most

---

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

notably with a 2003 report from the Health Effects Institute (HEI Accountability Working Group 2003), questions of “accountability” assessment — assessment of the extent to which regulatory actions taken to control air quality demonstrably affect health outcomes — have been propelled to the forefront of policy debates. A National Research Council report commissioned by the U.S. Congress recommended that an enhanced air quality management system should strive to take a more performance-oriented approach by tracking the effectiveness of specific control policies and creating accountability for results; similar calls for increased accountability have been echoed by others, including U.S. EPA (Hidy et al. 2011; Hubbell 2012; National Research Council 1992, 2004; U.S. EPA 2013). Increased emphasis on the direct study of the effectiveness of specific actions is an essential avenue to ensuring that epidemiological research continues to inform air quality control policies in the current regulatory climate.

## OVERVIEW

The goal of this report was to provide new analytic perspectives and statistical methods for what we refer to as direct-accountability assessment of the effectiveness of specific air quality regulatory interventions. Toward this end, we sharpened many of the distinctions surrounding accountability assessment initially raised by the HEI Accountability Working Group (2003) through discussion, development, and deployment of statistical methods for drawing causal inferences from observational data. The methods and analyses presented here are unified in their focus on anchoring accountability assessment in the estimation of the causal consequences of well-defined actions or interventions. These analytic perspectives are discussed in the context of two direct-accountability case studies pertaining to four different links in the chain of accountability, the related series of events leading from the intervention to the expected outcomes (See Preface; HEI Accountability Working Group 2003). The statistical methods described below consist of both established methods for drawing causal inferences from observational data and of our newly developed methodology for causal accountability assessment.

### CASE STUDY 1: PM<sub>10</sub> NONATTAINMENT DESIGNATIONS

The 1990 CAA amendments designated several counties or partial counties of the United States as being in nonattainment for PM<sub>10</sub>. These areas were designated as such (1) if the U.S. EPA had previously determined they were highly likely to violate the 1987 NAAQS for PM<sub>10</sub> or (2) if they had at least one monitored violation of the PM<sub>10</sub>

standard before January 1, 1989. A nonattainment designation required the state in which the area was located to develop a strategy to attain the standard by a target date and to revise its state implementation plan (SIP) accordingly. Counties not designated as nonattainment were not required to include new control strategies in SIP revisions pending additional monitoring. In Case Study 1, we present an analysis of the extent to which initial PM<sub>10</sub> nonattainment designations causally affected ambient PM<sub>10</sub> and health outcomes among Medicare beneficiaries. Our presentation of the case study focused on illustrating the most fundamental features of a causal-inference perspective on direct-accountability assessment. The analysis examined three links in the chain of accountability — regulatory action, ambient air quality, and human health response.

### CASE STUDY 2: SCRUBBER INSTALLATIONS ON COAL-FIRED POWER PLANTS

Even before the 1990 CAA amendments, various CAA regulations required or otherwise resulted in the installation of SO<sub>2</sub> scrubbers in new or existing EGUs. These included New Source Performance standards, Prevention of Significant Deterioration requirements for major sources, and implementation of the NAAQS for SO<sub>2</sub>. The 1990 CAA amendments added the Acid Rain Program (ARP), which instituted a requirement for major emissions reductions for both SO<sub>2</sub> and NO<sub>x</sub> from stationary pollution sources. One goal of this program was to reduce total SO<sub>2</sub> emissions by 10 million tons compared with the 1980 levels of 29.5 million tons per year. This reduction was to be achieved mostly through cutting emissions from EGUs. An integral strategy used to achieve the emissions-reduction goals of the program, especially among coal-fired EGUs, was the installation of flue-gas desulfurization equipment (“scrubbers”) to reduce SO<sub>2</sub> emissions. In Case Study 2, we present an analysis of the extent to which the presence of SO<sub>2</sub> scrubbers on coal-fired EGUs causally affected emissions of SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub>, as well as ambient concentrations of PM<sub>2.5</sub>. Our presentation of the case study focused on illustrating our newly developed statistical methods for multipollutant accountability assessment, which were designed to quantify the causal pathways through which a regulatory action affects ambient air quality. The results and limitations of the approach were interpreted against the vast backdrop of known relationships among scrubbers, emissions, and ambient PM<sub>2.5</sub>. The analysis examined three links in the chain of accountability — regulatory action, emissions, and ambient air quality.

### A NOTE ON THE WORD “CAUSAL”

Given the focus of this report, it is important to emphasize at the outset what the word “causal” means in the context of the statistical methods used here. Notions of causality are frequently viewed in air pollution epidemiology along a continuum of strength of evidence, in which causal evidence is the strongest for supporting conclusions about, say, biological mechanisms. As will be discussed in detail below, describing the methods in this report as “causal inference” methods does not refer to the strength of the evidence they supply, but rather to a framework for formalizing research questions in order to learn about the consequences of specific actions. A “cause” in this sense is an action (in this context, a particular intervention to control air quality) that might or might not occur, and an “effect” is a consequence of an explicitly defined cause. The statistical methods for causal inference used here first explicitly define causes and their effects and then use observational data to estimate the effects. The resulting estimates are not automatically guaranteed to have causal validity; rather, our framework provides a rigorous and principled way of clarifying and remedying some of the most common threats to validity that have plagued epidemiological studies. As such, the use of the term causal in this report should not be misconstrued as indicating any particular strength of evidence or degree of statistical significance. Causal estimates should be viewed as empirical estimates of well-defined consequences of specific actions. As in any epidemiological study, the estimates should be interpreted in light of the available data and the specifics of the statistical models used for estimation.

---

### SPECIFIC AIMS

---

1. Use a potential-outcomes framework to define causal effects of interest for single-pollutant accountability assessment and develop estimation methods. A causal-inference method for accountability research was developed that used principal stratification (Frangakis and Rubin 2002) to isolate the causal pathways leading from regulations to changes in air quality and health. The method allowed us to quantify and disentangle the causal effects of the regulation on health that were associated (1) with causal effects of the regulation on air quality and (2) with causal pathways capturing other factors that did not involve changes in air quality.
2. Define causal effects for multipollutant accountability assessment and develop estimation methods. Current statistical methods for assessing the consequences of air quality management rely on the specification of a

single pollutant and estimation of its health effects. We proposed a method for multipollutant accountability research of estimating the joint effect of a regulation on multiple pollutants, allowing estimation of the (possibly synergistic) downstream effects on health.

3. Develop national databases, conduct epidemiological studies, and disseminate software and results. We assembled and linked national data sets containing information on regulatory actions, emissions, ambient levels of criteria pollutants, health outcomes, and confounders for the entire United States. We then applied our proposed methods to the national data sets to estimate the impact on health indicators of specific regulations that targeted various pollutants. The necessary software and computational tools for our methods will be disseminated along with the results from the epidemiological studies.

---

### METHODS AND STUDY DESIGN

---

#### **PUBLICLY AVAILABLE DATA AND REPRODUCIBLE RESEARCH FOR ACCOUNTABILITY ASSESSMENT**

A key component of our development of methods for accountability assessment was the creation of data sources and tools for reproducible research. We have created a national, linked database containing information on ambient air quality monitoring, weather, population demographics, Medicare hospitalization and mortality outcomes, U.S. EPA nonattainment designations, continuous-emissions monitoring for more than 4000 EGUs in the United States, and emissions control technologies used at these units. Information in the database spans the years 1990 to 2015 and has unprecedented accuracy and granularity for conducting the types of accountability assessment discussed in this report.

Our efforts toward transparency and reproducibility have been focused on three areas. First, we developed R software programs to implement the newly developed methods; the programs will be made available on the Comprehensive R Archive Network. Second, we developed separate R software programs for the downloading, pre-processing, and linking of the data sources described below. With the exception of the Medicare data, all data used in this report are freely available and downloadable. The Medicare data are publicly available but must be purchased. The programs we are currently developing will allow anyone with R to automatically download and integrate the freely available data for use in their own research. Finally, we are working to make the specific data sets used

for our analyses (with appropriate privacy protections for the health data) accessible through the Harvard Dataverse, an online repository for sharing, citing, and preserving research data (Crosas 2011; King 2007). Distribution of our database and software will permit any study based on our database and software to be replicated.

### Ambient Monitoring Data

We developed scripts that retrieved daily and annual data at the monitor level from the U.S. EPA's Air Quality System (AQS) ([www2.epa.gov/aqs](http://www2.epa.gov/aqs)). These data, pre-processed by the U.S. EPA from hourly raw data, contain measurements corresponding to more than 10,000 monitors for the period 1990 to 2014 (approximately 5000 of which are currently active). The scripts allowed us to obtain data on criteria gases (ozone [O<sub>3</sub>], SO<sub>2</sub>, carbon monoxide [CO], and nitrogen dioxide [NO<sub>2</sub>]), particulate matter (PM<sub>2.5</sub>, non-Federal Reference Method PM<sub>2.5</sub>, PM<sub>10</sub> mass, and speciated PM<sub>2.5</sub>), lead, meteorological factors (wind speed, temperature, barometric pressure, and dew point), and toxics (hazardous air pollutants and volatile organic compounds).

### Medicare Health Outcomes Data

The Centers for Medicare and Medicaid Services (CMS) (<http://cms.hhs.gov/>) provide restricted-access daily Medicare health outcomes data at the zip-code level from 1999 to 2015, including data on the following:

- Cohort: All Medicare enrollees by year of enrollment, including age, gender, race, state, and five- and nine-digit zip-code identifiers for their residence (40,000,000 people per year).
- Mortality: Date of death for enrollees in the Medicare cohort.

- Hospitalizations: Hospitalization records for all Medicare enrollees, including date of hospitalization, length of stay, International Classification of Diseases (ICD) primary and secondary diagnostic and procedure codes associated with the hospitalization, and the costs billed to Medicare for the hospitalization (for details see Domini et al. 2006; Peng et al. 2008; Zanobetti et al. 2014).

### Regulatory Data Sources

County-level nonattainment designations are available from the U.S. EPA Green Book for all criteria pollutants since 1978 ([http://www.epa.gov/airquality/greenbook/data\\_download.html](http://www.epa.gov/airquality/greenbook/data_download.html)). For the period 1995–2012, open-access daily data are available at the EGU and power-plant level from the U.S. EPA's Air Markets Program Data (AMPD) (<http://ampd.epa.gov/ampd/>), where a power plant is defined as a facility with one or more EGUs. We collected data on 4164 EGUs belonging to 1248 facilities, the totality of facilities that participated in the ARP. These monitored facilities (or “ARP plants”) are an important source of SO<sub>2</sub> emissions. They represent approximately 20% of all power plants in the United States and, yet, as shown in Figure 1, account for 75% of fuel-combustion emissions and 65% of overall emissions.

For each unit, we collected information about location (e.g., state, county, latitude, and longitude); ARP phase (I, II, opt-in, substitution, and compensating); SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub> emissions from continuous-emissions monitors (CEMs); average NO<sub>x</sub> emissions rates; heat input, gross load, steam load, operating time, and status; primary and secondary fuel types (e.g., coal, diesel oil, and natural gas); and scrubber technologies (i.e., whether a scrubber is installed and, if so, the technology it used) for SO<sub>2</sub>, NO<sub>x</sub>, and PM.

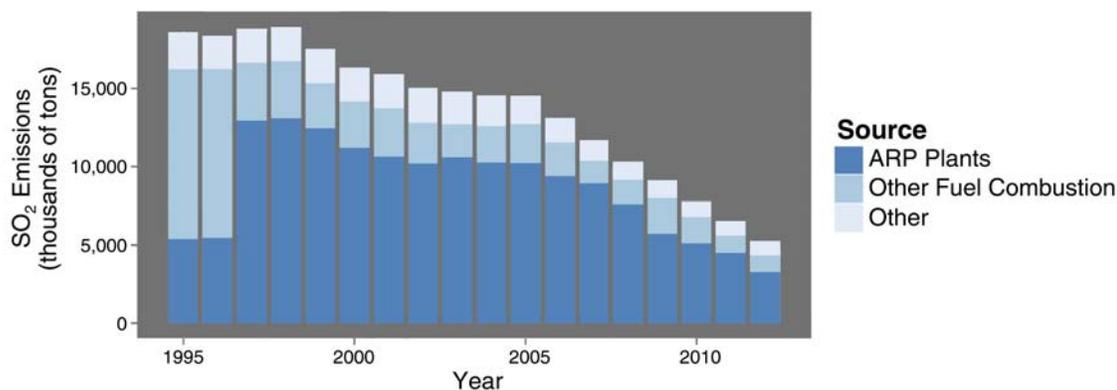


Figure 1. Relative importance of SO<sub>2</sub> emissions from ARP-monitored facilities.

### Other Supporting Data Sources

We obtained demographic information for the year 2000 from the U.S. Census Bureau (<http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>), annual temperature data from the National Climatic Data Center ([www.ncdc.noaa.gov/cdo-web/](http://www.ncdc.noaa.gov/cdo-web/)), and county-level smoking rates from the Centers for Disease Control Behavioral Risk Factor Surveillance System ([www.cdc.gov/brfss/data\\_tools.htm](http://www.cdc.gov/brfss/data_tools.htm)). From the U.S. Energy Information Administration (EIA) ([www.eia.gov/](http://www.eia.gov/)), we obtained monthly and annual data (from forms EIA-767, EIA-906A, and EIA-923) for the period 1985–2012 at the power-plant level, including the year that each unit was or is expected to be in compliance, the strategy for PM compliance, actual or projected in-service and retirement dates, primary fuels and alternate fuel capacity, monthly net electrical generation, fuel, and monthly heat, ash, and sulfur contents. We also merged EIA information with that in the AMPD database, relying on the unique facility and unit identifiers that correspond unambiguously to power-generating units.

### Linking Existing Data

Strategies to link data had to take into account that the data sources that we used report on data from various geographical units, and are thus spatially misaligned. For example, ARP and AQS data are coordinate-based, whereas Medicare enrollee data is available at the zip-code level.

**AQS and Medicare** In order to link ambient monitoring sites to zip codes, which was necessary to link AQS data to Medicare health outcomes, we identified the centroid of each zip code in the United States and enumerated all EPA monitoring sites within a certain distance. When a zip code was close to more than one monitor, that zip code was linked only to the closest monitor. After each zip code was linked to at most one monitoring site, Medicare data were then obtained at the zip code and aggregated to the level of the monitoring site by combining data on all zip codes assigned to each monitor. Our process had several advantages: it ensured that each zip code was reliably assigned to only one monitoring site, was computationally efficient, was customizable (e.g., could be used to link zip codes at various distances), and decoupled the zip code–AQS linkage from the linkage with Medicare data so that the linkage could be conducted without any specialized knowledge of, or access to, Medicare data.

**AMPD and Medicare and AMPD and AQS** Our linkage algorithm could also be used to link AMPD power plants to AQS monitoring sites or zip codes for our analyses of the ARP. Our goal was to provide computer programs and source

code for the monitor–zip code linkage that could be re-implemented for any specified set of monitoring locations (e.g., all PM<sub>10</sub> monitoring sites with “population-oriented” monitors or all monitoring sites containing monitors for both PM<sub>10</sub> and O<sub>3</sub>). Another data process improvement pertained to inclusion and exclusion criteria for monitoring sites; for example, we were able to reliably exclude monitors that were calibrated on a micro-scale or those for which there were not enough daily measurements for a given year.

### The Harvard Dataverse

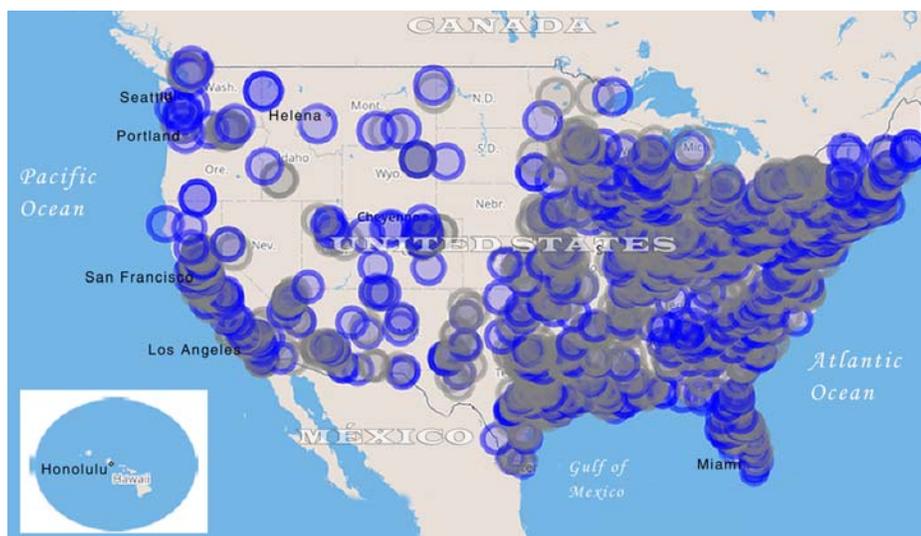
Our data sets and databases (linking EPA and EIA sources and simulated Medicare information) are accessible from a public access Harvard Dataverse (<https://dataverse.harvard.edu/dataverse/airqualregs>), an online repository for sharing, citing, and preserving research data (see King 2007).

### AREPA R Package

We have developed software programs within the R statistical environment to allow for easier and broader distribution of our results. In particular, we are developing the AREPA (An R package for EPA, which for the purposes of EPA data retrieving and processing) package in a private GitHub repository (<https://github.com/czigler/arepa>). Our codebase leverages the wealth of tools provided by R, and more specifically uses the fast and efficient in-memory big data manipulation with `data.table` (<http://cran.r-project.org/web/packages/data.table/index.html>) and the geographic-information system capacities with `sp` (<http://cran.r-project.org/web/packages/sp/index.html>). The AREPA repository will become public and the package of software programs will be made available on the Comprehensive R Archive Network at <http://cran.r-project.org/>.

The AREPA package is currently used within our group and provides three main groups of functionalities to improve the efficiency and reproducibility of our workflow:

- Script-based downloads for daily and annual AQS data (as described above in the section on ambient monitoring data).
- Spatial linkage procedures that implement the methods described above in the Methods section on existing data linkage. For example, Figure 2 illustrates the spatial linkage between AMPD power plants and AQS PM<sub>2.5</sub> monitors in 2010 with a default radius of 100 km.
- An indexed data set from which Medicare data can be retrieved at the zip-code level and then aggregated around AQS monitors or AMPD power plants. We will provide simulated Medicare data to illustrate the data formats used in our routines so that they can be used by other research groups.



**Figure 2. Spatial linkage between AMPD power plants and AQS PM<sub>2.5</sub> monitors in 2010.** Blue circles correspond to a 100-km radius around a power plant successfully linked to a unique monitor; gray circles correspond to power plants that have been discarded for lack of a unique monitor within the 100-km range.

## STATISTICAL PERSPECTIVES FOR CAUSAL ACCOUNTABILITY ASSESSMENT

The role of causality is of obvious import for informing policy decisions, and the causal validity (or lack thereof) of epidemiological evidence has always been central to the integration of scientific evidence into policy recommendations (U.S. EPA 2009). However, approaches to infer causality from available observational data can vary depending on the scientific question of interest and the data available for analysis.

Causal inference in air pollution epidemiology has most commonly been undertaken within a classical paradigm, which considers causal validity on a continuum according to how likely an observed association (e.g., between pollution and health) can be interpreted as causal (Glass et al. 2013). This continuum is explicitly considered in the approach to the Integrated Science Assessments conducted by the EPA, which classify evidence of the association between pollution exposure and health as a “causal relationship,” “likely to be a causal relationship,” “suggestive of a causal relationship,” “inadequate to infer a causal relationship,” or “not likely to be a causal relationship.” Even in the absence of the word “causal,” the bulk of air pollution epidemiology has been implicitly undertaken with this classical approach; an exposure–response relationship between pollution and health is estimated (e.g., in a cohort study), then a judgment is made as to whether this relationship can be reasonably interpreted as causal, and finally, hypothetical changes in exposure are input into the

exposure–response function to infer the resulting “health effect” that would be caused by such a change in pollution. Many such studies have been integral to issues of accountability (Correia et al. 2013; Dockery et al. 1993; Laden et al. 2006; Pope et al. 2009; Zeger et al. 2008). Although the strength of evidence needed for policy action may vary with the specific context, establishment of exposure–response relationships within the classical paradigm have provided support for a variety of air quality control policies.

As an alternative to the classical paradigm, the potential-outcomes paradigm for causal inference has the distinctive feature that causal effects are explicitly defined as consequences of specific actions (Rubin 1978). Rather than infer causality based on belief of whether an estimated exposure–response relationship can be interpreted as causal, potential-outcomes methods entail definition of a clearly defined action (a cause), the effects of which are of interest. Some existing accountability assessments have been (often implicitly) undertaken within a potential-outcomes paradigm for causal inference, sometimes framed as intervention studies that analyze a large, abrupt change in air pollution. The common thread in these lines of research is the application of the core tenets of experimentation to observational settings (Chay et al. 2003; Chen et al. 2013; Clancy et al. 2002; Currie and Walker 2011; Deschenes et al. 2012; Friedman et al. 2001; Greenstone 2004; Hedley et al. 2002; Moore et al. 2010; Pope et al. 2007; Rich et al. 2012; Tonne et al. 2008; Zigler et al. 2012).

### Direct versus Indirect Accountability

Studies framed as accountability studies can be classified according to the specific scientific question of interest. Studies that answered questions in the form of “What is the relationship between exposure to pollution and health outcomes?” can aptly be described as indirect-accountability studies (Zigler and Dominici 2014). This type of question has been at the center of air pollution epidemiology for decades, and answers typically came in the form of exposure–response relationships between (changes in) pollution exposure and (changes in) health outcomes. Importantly, these studies did not consider the effectiveness of any specific regulatory action, but provided valuable evidence for indirectly predicting the impact of policies. For example, U.S. EPA has routinely used exposure–response estimates to estimate the expected benefits of current and future policies; if a policy reduced (or was expected to reduce) pollution by a certain amount, then the exposure–response relationship indirectly implied the health impact of the policy insofar as the relationship had been deemed causal (U.S. EPA 2009, 2010, 2012). Indirect-accountability assessments typically assume that any observed exposure–response relationship would persist amid the complex realities of actual regulatory implementation that in reality typically affect a variety of factors. As a consequence, health impacts of regulatory interventions may not be accurately characterized by indirectly applying exposure–response estimates to accountability assessments.

We focused on a slightly different perspective on accountability assessment that we term direct accountability (Zigler and Dominici 2014). Direct-accountability studies target a different scientific question than studies of exposure–response relationships. Rather than investigate the relationship between pollution and health, these direct-accountability studies answer the question “What is the relationship between a specific regulatory intervention and health?” These studies are “direct” accountability studies in that they directly evaluate the effectiveness of well-defined regulatory actions; the resulting evidence informs questions as to the actual health benefits of these actions. We argue that direct-accountability assessments are best equipped to meet the demands of shifting legislative, judicial, and political environments fraught with questions surrounding the effectiveness of specific policies. Of particular importance is the noted lack of direct evaluations of broad, complex regulatory interventions, which are of utmost relevance to policy debates (Health Effects Institute 2010; van Erp et al. 2012).

The analytic perspectives and statistical methods described in this report, namely, those rooted in potential-outcomes methods for causal inference, are particularly well suited to answering questions of direct accountability. The purpose for distinguishing between direct and

indirect accountability is not to highlight the need for causal versus associational evidence, as all research to provide such evidence shares the goal of establishing causality. Rather, we argue that today’s regulators and policymakers would be better informed by evidence of the effectiveness of specific control policies, and such evidence for previously implemented policies can augment epidemiological approaches tailored to exposure–response estimation. In an environment of skepticism and doubt about results drawn from observational data, the analysis of the consequences of specific interventions using approaches rooted in potential-outcomes thinking can clarify the basis for drawing causal inferences and bring a higher level of credibility to evidence used to support policy decisions (Dominici et al. 2014; Zigler and Dominici 2014).

### Potential-Outcomes Methods: Framing Observational Studies as Hypothetical (Approximate) Randomized Experiments

The underlying features of randomized studies that make them the “gold standard” for generating causal evidence remain pertinent to causal accountability assessment; potential-outcomes methods frame observational studies according to how well they can approximate randomized experiments (Hernan et al. 2008; Rubin 2008). The key idea is to define an experiment (possibly hypothetical) as consisting of an “intervention condition” and a “control condition,” such that if populations could be randomly assigned to these conditions, differences in observed health outcomes between the conditions could be interpreted as causal effects of the intervention. Although defining the intervention condition in accountability studies can be straightforward (e.g., it will likely be a regulatory action that actually occurred), framing an accountability study as a hypothetical experiment forces the specification of some alternative action that might have otherwise occurred to serve as a relevant control condition. This exercise formalizes the research question by explicitly defining a causal effect as a comparison between what would happen under well-defined competing conditions. Hence the name of the potential-outcomes paradigm: a causal effect of Action A relative to Action B is defined as the comparison of the potential outcome if Action A were taken with the potential outcome if Action B were taken. Thus, the salient question for accountability is not “Did health outcomes change after the intervention?” but rather “Are health outcomes different after the intervention than they would have been under a specific alternative action?” For example, our first case study (Accountability Assessment of PM<sub>10</sub> Nonattainment Designations in the Western United States) attempts to answer the question: “Are Medicare health outcomes in

PM<sub>10</sub> nonattainment areas different from what they would have been if the nonattainment designations had never occurred?” Using the language of experimental design, the hypothetical “intervention” condition is the set of observed PM<sub>10</sub> nonattainment designations, whereas the hypothetical control condition is the set of areas that never received such a designation.

Of utmost importance is that the definition of the causal effect of interest be purely conceptual and made explicit without regard to any assumed statistical model. Different models could be used to actually estimate this effect, but the effect itself, along with its interpretation, must remain consistent regardless of the modeling approach. This clarity is essential for producing policy-relevant evidence. Traditional epidemiological studies, by contrast, (1) do not necessarily explicate an action defining effects of interest and (2) define health effects with parameters (e.g., regression coefficients) in a statistical model.

The fundamental problem of estimating causal effects with comparisons between potential outcomes under competing intervention and control conditions is that if the intervention is enacted, the outcomes under the control condition are unobserved. For example, evaluating the effect of a PM<sub>10</sub> nonattainment designation requires knowledge of what would have potentially happened if the nonattainment designations had never occurred. Hypothetical scenarios that never actually occurred are often referred to as “counterfactual” scenarios, and estimating what would have happened under such scenarios is perhaps the most important challenge for direct-accountability assessments.

Counterfactual scenarios have been explicitly considered, for example, in U.S. EPA cost–benefit analyses of the CAA mandated by Section 812 of the act, which project two counterfactual pollution scenarios: one scenario assumes past exposure patterns would have continued without the 1990 CAA Amendments and the second scenario assumes an expected change in exposure patterns under full implementation of the 1990 Amendments. These two scenarios are coupled with exposure–response functions from the epidemiological literature to project counterfactual health scenarios that form the basis of the health-benefits analyses (U.S. EPA 2010, 2012). However, these counterfactual scenarios are not validated against studies of actual interventions, and thus are not sufficient for fully characterizing the relationships between regulatory strategies and health (HEI Accountability Working Group 2003).

Rather than project counterfactual scenarios by combining assumed exposure patterns with exposure–response estimates, potential-outcomes approaches

typically use actual data from the “control group” of the hypothetical experiment to learn what could have happened under the hypothetical control condition, rendering identification of an appropriate control population of vital importance. When assessing the impact of regulatory intervention in comparison to what would have happened absent the intervention, control populations can be defined based on time (e.g., a population before promulgation of a regulation) or space (e.g., if some areas are subject to an intervention and others not). Whether outcomes in the control population can actually characterize what would have occurred absent the intervention boils down to the familiar concept of confounding, although what constitutes a confounder of the effect of an intervention is slightly different from the common conception of a confounder as something that is related to both pollution exposure and health.

For direct accountability studies, a comparison of outcomes of the intervention and of control conditions is unconfounded if the two populations are comparable for factors that relate to outcomes. An unconfounded comparison of outcomes of the intervention and control conditions yields an estimate of the causal effect. If the two populations differ on important factors related to outcomes, then such a comparison is a convolution of differences related to the intervention and to other factors. Thus, if an important factor relating to health (e.g., smoking behavior) is comparable across the intervention and control populations, then the factor (smoking behavior) is not a confounder in the assessment of the intervention. In the typical exposure–response studies, a confounder is generally regarded as a factor that is simultaneously associated with pollution exposure and health outcomes. In both types of studies, the definition of a confounder is a factor that is associated with exposure and outcome. The key difference in a direct-accountability study is that the exposure is actually the intervention.

### Methods for Confounding Adjustment: Propensity Scores

There are a variety of analytic tools available to address confounding in nonrandomized accountability studies. Specialized study designs, often described as quasi-experiments, circumvent the need to consider confounding directly, as they support assumptions that an intervention was quasi-randomized in the sense that it is unrelated to health outcomes (Dominici et al. 2014; Greenstone and Gayer 2009). Absent the availability of such specialized circumstances, methods for confounding adjustment (e.g., matching, weighting, stratification, or standardization) adjust for differences between intervention and control populations so that comparison groups

can be regarded as similar on the basis of observed factors, thus mimicking the design of a randomized study.

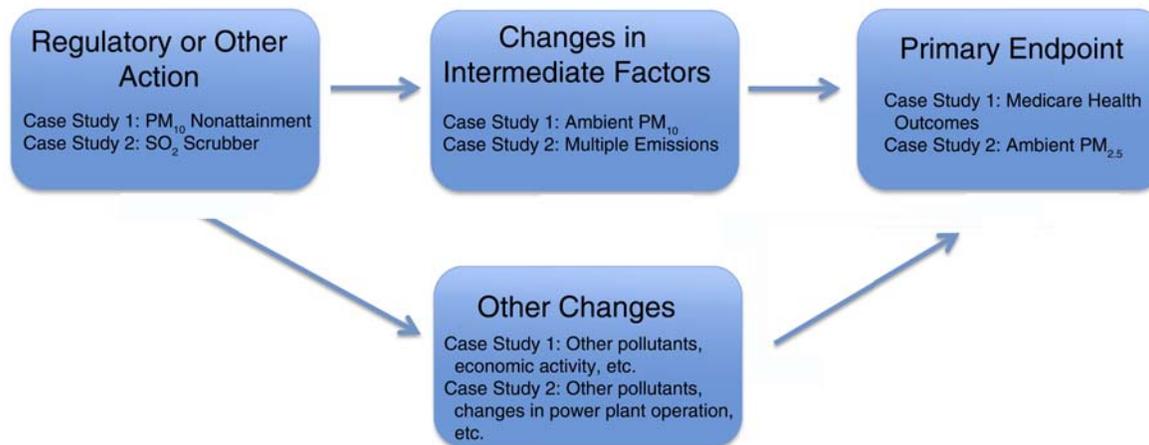
One broad class of methods for confounding adjustment relies on the propensity score (Robins et al. 2000; Rosenbaum and Rubin 1983; Rubin 2008; Stuart 2010). Propensity score methods share the same objective as, say, adjusting for covariates in a regression model, but have been shown to have several benefits in comparison to reliance on parametric regression models. Propensity scores represent a dimension-reduction procedure in which multivariate confounding information is condensed into a one-number confounder summary called the estimated propensity score. Observations with similar estimated propensity scores can be regarded as being similar on the basis of all of the covariates that were used for its estimation. The value of the estimated propensity score can then be used to adjust for confounding via matching, weighting, or subclassifying the observed sample based on the estimated propensity score in order to ensure that groups of treated and control observations are comparable. If the propensity score model is adequately specified (i.e., if there is no unmeasured confounding), then average outcomes in control observations represent what would have happened (on average) in treatment observations with similar values of the propensity score, as would be the case in a randomized study. In the analysis of Case Study 1, we used propensity scores for confounding adjustment and we discuss additional considerations related to confounding in the context of the analysis.

**Causal Pathways Analyses: Causal Mediation Analysis and Principal Stratification**

An objective of causal accountability assessment is to quantify the relative importance of the possible causal pathways that constitute links in the chain of accountability. For example, one set of questions may relate to the extent to which the causal effect of an intervention on health outcomes acts through reducing ambient pollution, or there may be questions about the extent to which an intervention effect on ambient pollution is mediated through changes in specific emissions. Figure 3 presents a schematic representation of causal pathways for accountability assessment.

Understanding the pathways through which an intervention affects ambient air quality or health outcomes is critical for informing policy decisions. From this perspective, intermediate factors in the chain of accountability that lie between the regulatory action and human health response can be regarded as lying “on the causal pathway.” Because such intermediate outcomes are posttreatment, concomitant variables that are expected to simultaneously be affected by the intervention and have bearing on outcomes, standard regression adjustments will not permit estimation of causal effects (Rosenbaum 1984). We considered two related causal frameworks for characterizing causal effects with intermediate variables: causal mediation analysis and principal stratification.

Causal mediation analysis is a framework designed to isolate specific causal pathways in order to assess whether



**Figure 3. Schematic description of direct and indirect causal pathways for accountability assessment.** Air quality interventions are typically intended to impact primary pollution and health outcomes through reducing specific emissions and/or ambient pollutants (indirect effects) but can, in reality, impact outcomes through other causal pathways (direct effects).

the causal effect of an intervention on an outcome is mediated through the causal effect of the intervention on an intermediate variable (VanderWeele and Vansteelandt 2009). Causal mediation analysis implies two hypothetical interventions in the context of accountability. The first hypothetical intervention represents the treatment and control groups (described above). The second hypothetical intervention is defined as one that acts directly on the intermediate variable, independently of the first intervention. In Case Study 2, the first hypothetical intervention represents the presence of an SO<sub>2</sub> scrubber on a coal-fired power plant, which is the actual regulatory intervention of interest. The second intervention is purely hypothetical and corresponds to a way in which power-plant emissions could be manipulated independently of a scrubber. Definition of potential outcomes based on these two interventions permits decomposition of the total causal effect of an intervention into effects that are direct and indirect effects (Robins and Greenland 1992). In the context of Case Study 2, a direct effect corresponds to the causal effect of the scrubber that acts directly on ambient pollution in that it is attributable to causal pathways not involving emissions; an indirect effect is the effect of a scrubber on ambient pollution that can be attributed to causal emissions reductions. Causal mediation analysis is explained and illustrated with more technical detail below in the discussion of Case Study 2 in the Results section.

Principal stratification is a related framework for causal inference with intermediate variables, but considers only one hypothetical intervention (i.e., the regulatory intervention being assessed) (Frangakis and Rubin 2002). In the context of accountability assessment, principal stratification aims to quantify the extent to which the causal effects of an intervention on the primary outcome coincide with the causal effects of the intervention on the intermediate variable (Zigler et al. 2012). For example, in Case Study 1, we investigated the extent to which a causal effect of PM<sub>10</sub> nonattainment designation on Medicare health outcomes coincided with the causal effect of the designation on ambient pollution. Towards this end, using principal stratification, we defined two types of causal effects: dissociative and associative causal effects. The dissociative effects quantified the extent to which the intervention causally affected health outcomes when the intervention did not causally affect pollution. Dissociative effects are similar to direct effects in the mediation analysis framework in that they represent causal health effects of an intervention indicative of causal pathways other than ambient pollution. Associative effects quantify the causal effect of the intervention on health when the intervention causally affects ambient pollution. Associative effects are similar to (but distinct from) indirect effects in the causal mediation

analysis framework. An associative effect that is larger than the dissociative effect indicates that the causal effect of the intervention on health outcomes is greater in areas where pollution was causally affected than in areas where there is little or no effect of the intervention on pollution. Such a finding would suggest the presence of a causal pathway whereby the intervention affects health through changing pollution. Dissociative effects that are similar in size to associative effects indicate that the health impact of the intervention is similar regardless of whether the intervention causally affected ambient pollution, which suggests the presence of other causal pathways through which the intervention affects health without changing pollution.

The theoretical and technical differences between principal stratification and causal mediation analysis have been closely examined in the causal inference literature (e.g., Joffe and Greene 2009; Pearl 2011; Rubin 2004; VanderWeele 2008) in the setting of a single mediating factor. Our analysis of Case Study 1 illustrates the use of principal stratification in the analysis of PM<sub>10</sub> nonattainment designations. Our analysis of Case Study 2 provides technical details of new methods for both principal stratification and causal mediation analysis in the multipollutant context and interprets both analyses in the context of this particular case study.

---

## RESULTS

---

### CASE STUDY 1: ACCOUNTABILITY ASSESSMENT OF PM<sub>10</sub> NONATTAINMENT DESIGNATIONS IN THE WESTERN UNITED STATES

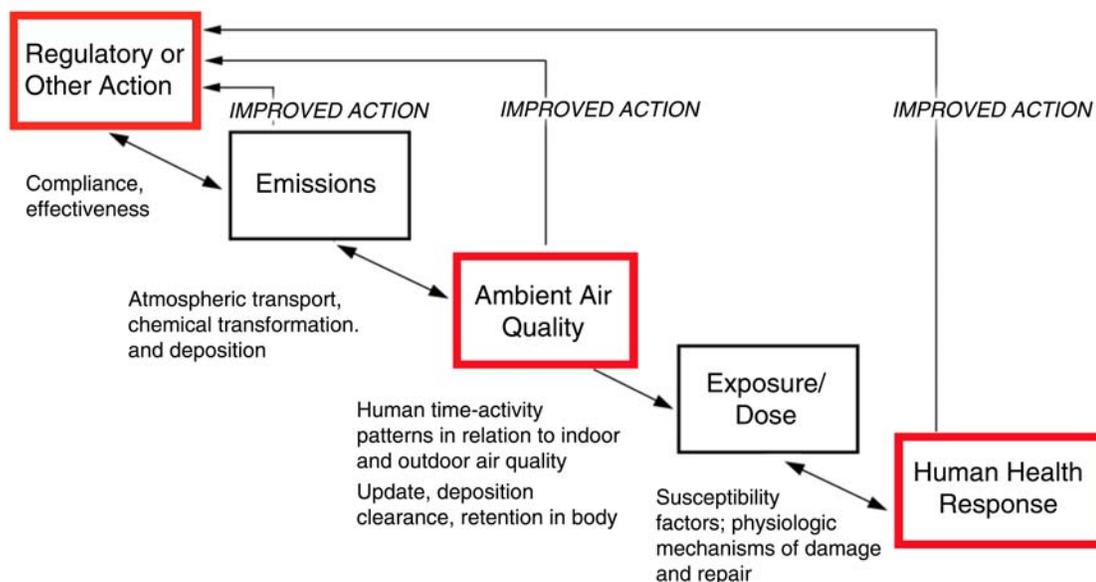
In our first case study, we employed the analytical perspectives outlined in the previous section to provide the first direct-accountability assessment of the health impacts of one integral regulatory strategy defined under the 1990 amendments to the CAA — the initial designation of areas as nonattainment with the 1987 NAAQS for PM<sub>10</sub>. In contrast to earlier efforts to examine the impact of the CAA as a whole (HEI Accountability Working Group 2003; U.S. EPA 2010), we focused our analysis on the initial PM<sub>10</sub> nonattainment designations. We chose this focus for two important reasons: (1) the decision process whereby the U.S. EPA sets NAAQS and initiates nonattainment designations is one integral tool for managing air quality under the CAA, thus quantifying the effects of this specific process can provide valuable evidence to support future adaptations to the process and (2) a focused characterization of the health impact of a specific set of regulatory decisions provides more targeted direct accountability of

the effectiveness of a specific regulatory decision, which in turn yields more specific evidence for informing future policies than would earlier assessments that considered the whole of the CAA.

Our analysis differed from traditional epidemiological investigations of the long-term association between pollution exposure and health in that we adopted an analytical perspective designed specifically to estimate causal effects of a specific set of actions (i.e., the nonattainment designations) rather than to characterize the exposure–response relationship between pollution and health in a time frame that included myriad regulatory actions that contributed to improved air quality. Specifically, we used a principled causal-inference framework to assess whether the initial PM<sub>10</sub> nonattainment designations caused improvements in Medicare health outcomes. In accordance with the chain of accountability, we viewed changes in ambient PM<sub>10</sub> as intermediates on the causal pathway between regulatory decisions and health outcomes, representing three key links of the chain of accountability as depicted in Figure 4. (HEI Accountability Working Group 2003). In addition to estimating overall effects of nonattainment designations on ambient PM<sub>10</sub> and Medicare health outcomes, our approach provided additional information about the relative importance of different causal pathways through which regulatory decisions may affect health.

**Linked Data Sources**

We assembled a national, linked database using the tools described earlier in the Methods section to conduct our investigation. The study population consisted of U.S. Medicare beneficiaries living within 6 miles of a PM<sub>10</sub> monitoring location in 2001. The locations used were U.S. EPA monitoring stations located in the Western United States that were in operation at any point between 1990 and 2001. This region was chosen because virtually all initial nonattainment designations for PM<sub>10</sub> occurred in this part of the country. From the U.S. EPA Green Book, we enumerated every county in the United States designated as nonattainment for PM<sub>10</sub> between 1990 and 1995. Annual average ambient PM<sub>10</sub> concentrations from 1990–2001 were obtained from pollution monitor locations in the AQS database. Annual average PM<sub>10</sub> concentrations were regarded as missing if the percentage of valid measurements was less than 67%. Health data were assembled from CMS Medicare Part A hospital claims and enrollment data. From the CMS enrollment file, we enumerated all Medicare beneficiaries residing in a zip code within 6 miles of a pollution monitor during 2001. Beneficiaries living within 6 miles of multiple monitors were linked to the monitor closest to their zip code of residence. Data available on Medicare beneficiaries included basic demographic information, mortality information, and hospitalization records.



**Figure 4. Links in the chain of accountability considered in Case Study 1.** In our analysis Regulatory or Other Action refers to the initial PM<sub>10</sub> nonattainment designations in 1990–1995, Ambient Air Quality to the annual average ambient PM<sub>10</sub> concentration in 1999–2001, and Human Health Response to the Medicare health outcomes. (Adapted from Figure 1, Health Effects Institute 2010.)

Hospital billing claims data were used to identify hospitalizations for illnesses related to cardiovascular disease (CVD) and to the respiratory system. CVD-related hospitalizations were defined as those having ICD9-clinical modification (CM) codes 390.xx to 495.xx. Hospitalizations related to the respiratory system were defined as those relating to chronic obstructive pulmonary disease (ICD9-CM 490.xx to 492.xx) or respiratory tract infections (ICD9-CM 464.xx to 466.xx and 480.xx to 487.xx). Individual-level health data were aggregated to the level of the monitoring location to yield average demographic information (average age, percentage female, etc.) and outcome rates (mortality and hospitalization rates) for all beneficiaries living within 6 miles of each monitoring location.

We augmented the pollution–health linked database with county-level information from the 2000 U.S. Census and from the Center for Disease Control and Prevention Behavioral Risk Factor Surveillance Survey. County-level information includes population demographics, urbanicity, and smoking rates. Additionally, county-level values of annual average daily maximum temperature in 1990 were obtained by averaging across monitoring stations available from the National Climatic Data Center. Table 1 summarizes the linked information obtained in the database. For our analysis, data were considered at the monitor level, that is, for each monitoring location we have a specific location (latitude and longitude), measures of ambient pollution, demographic characteristics of the county

**Table 1.** Summary Statistics for Covariates and Outcomes Available for the Analysis of PM<sub>10</sub> Nonattainment Designations<sup>a</sup>

	Attainment Areas		Nonattainment Areas	
	Mean	SD	Mean	SD
<b>Monitor Data</b>				
Ambient PM <sub>10</sub> in 1990	26.36	6.55	39.2	12.75
<b>Medicare Data</b>				
Medicare beneficiaries*	5,813.33	9,579.5	8,767.5	13,387.71
Age*	74.77	1.1	74.7	1.31
Female* (%)	54.81	4.61	55.53	5.56
White* (%)	90.29	13.66	86.79	13.51
Black* (%)	1.48	3.28	3.14	5.14
<b>County-Level Data</b>				
Population*	889,937.31	1,472,176.42	3,380,578.87	5,125,807.84
Housing density*	0.42	0.08	0.4	0.08
Urban living* (%)	71.64	23.25	84.95	18.89
Median income*	42,148.87	10,415.55	40,873.47	7,764.82
High school graduates* (%)	83.8	6.23	79.9	8.52
5-Year migration rate*	22.73	6.22	20.52	6.26
Smoking rate*	18.5	5.51	19.61	4.32
Annual maximum temperature*	65.75	6.6	72.39	10.14
Hispanic (%)	16.13	13.93	24.81	19.51
White (%)	73.22	18.22	63.56	20.77
Black (%)	1.88	2.76	3.74	3.29
Female (%)	49.97	1.31	49.99	1.38
<b>Pollution and Health Outcomes Variables</b>				
Ambient PM <sub>10</sub> 1999-2001	21.58	6.43	31.56	13.28
Mortality rate: all cause	62.58	16.95	62.51	12.4
Hospitalization rate: CVD	83.74	24.24	92.09	26.65
Hospitalization rate: respiratory	28.39	17.05	28.41	12.78

<sup>a</sup> Variables marked with \* are those included in the model that estimates the propensity score and are used for additional covariate adjustment in models for pollution and health outcomes.

containing the monitor, and aggregated health information on all Medicare beneficiaries residing within a 6-mile radius. The initial analysis data set contained the 547 monitoring locations depicted in both panels of Figure 5, with health data comprised of information on 3,971,610 Medicare beneficiaries. Among these 547 locations, 268 are located in nonattainment areas, corresponding to 2,349,691 Medicare beneficiaries. Note that monitoring locations with fewer than 20 Medicare beneficiaries residing within the 6-mile radius in 2001 were excluded.

The outcome variables for our analysis were: (1) the average annual ambient concentration of  $PM_{10}$  during 1999–2001; (2) the all-cause mortality rate (number of deaths per 1000 beneficiaries); (3) the CVD-related hospitalization rate (number of hospitalizations per person-year); and (4) the respiratory-related hospitalization rate (number of hospitalizations per person-year). Person-years were used in the analysis of the hospitalization outcomes to account for the fact that beneficiaries can die or unenroll from Medicare during the year, and hospitalization records were only available during the time period of enrollment. In contrast, mortality was known regardless of Medicare enrollment status. All other variables listed in

Table 1 were considered covariates in our analysis. Note that some covariate values were measured after the nonattainment designations: census variables were from the 2000 Census and Medicare demographic data were measured in 2001. We assumed that such variables were not affected by the nonattainment designations and as such were reliable proxies for the same quantities in the years preceding the nonattainment designations.

Our analysis was confronted with two types of missing ambient pollution data. First, 284 monitoring locations (131 in nonattainment areas) had missing  $PM_{10}$  measurements in 1990. Missing values of average annual ambient pollution in 1990 were singly imputed using posterior mean predictions from a spatial hierarchical random effects model as described in section A.1 of Appendix A (available on the HEI Web site). The second group of missing pollution data was missing average annual ambient concentrations during 1999–2001. Average ambient  $PM_{10}$  concentrations for 1999–2001 were missing for 157 monitoring locations (70 in nonattainment areas). These follow-up pollution measures were used as outcomes in our analysis and were multiply imputed as a byproduct of our Bayesian estimation procedure in the models used to estimate causal effects (described below).

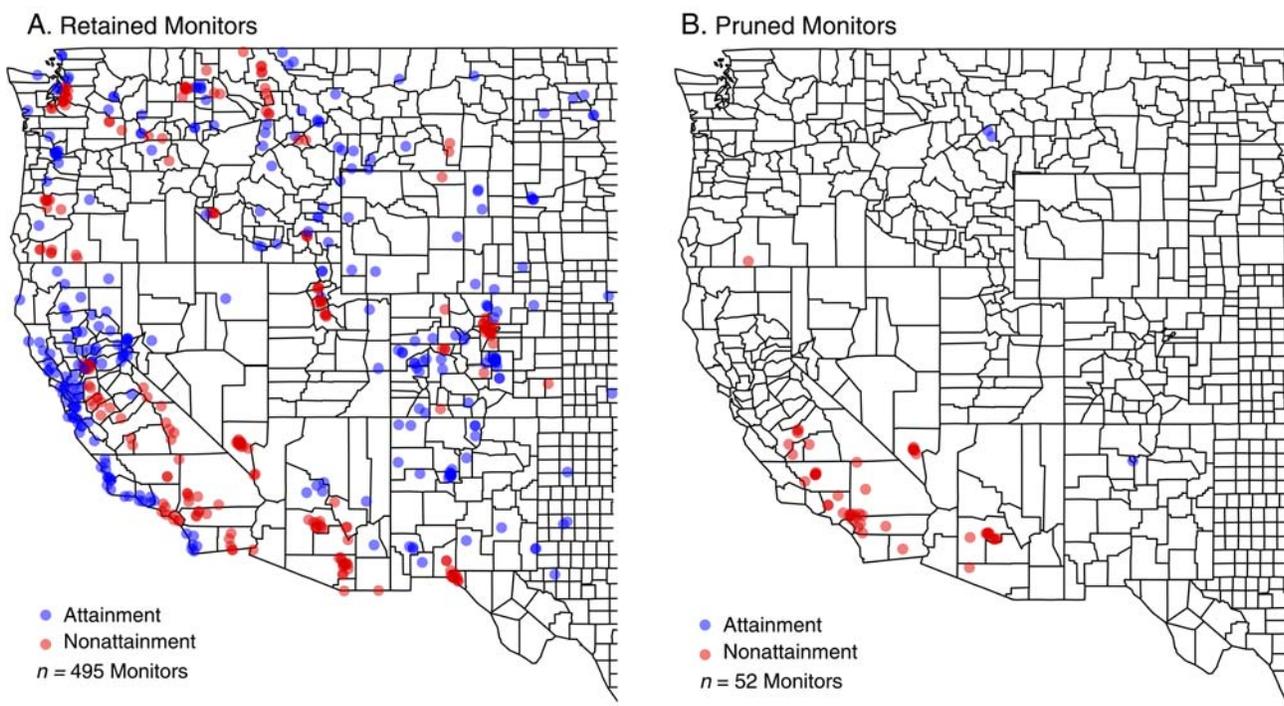


Figure 5. Locations of all 547  $PM_{10}$  monitoring locations available for analysis. (Left) The 495 locations retained after pruning by application of propensity scores. (Right) Locations pruned because there was no propensity score overlap.

Missing values of county-level average annual daily maximum temperature were imputed for 61 monitoring locations (12 in nonattainment areas) using  $k$  nearest-neighbor mean imputation. All other covariates and outcomes listed in Table 1 were fully observed.

### **Defining the Intervention for Direct-Accountability Assessment: Initial PM<sub>10</sub> Nonattainment Designations**

When the 1990 amendments to the CAA took effect, areas (counties or partial counties) in United States that were observed to be or had a high probability of being in violation of the 1987 NAAQS for PM<sub>10</sub> were designated as being in nonattainment of the standard. All other areas of the country were designated as unclassifiable for PM<sub>10</sub>. In the Western United States, some initially unclassifiable counties were subsequently designated as nonattainment between 1990–1995, based on increased availability of monitoring data. A nonattainment designation required the state containing that county to develop a strategy to attain the standards by a target date and to revise the SIP accordingly. Counties not designated as nonattainment were not required to include new control strategies in SIP revisions pending additional monitoring. Target dates for attainment of the standard varied by area within the regulatory regime, but the intent of the Act was that most nonattainment areas should achieve the standard by 2001. For the purposes of our analysis, we considered the “initial” nonattainment designations to consist of any such designation that occurred between 1990 and 1995, because the differences in features for areas designated in 1990 and those areas designated in the next few years are thought to relate more to procedural issues and availability of data than to air quality per se. The “intervention” for this analysis was these initial nonattainment designations, which either occurred or did not occur in every area of the United States.

### **Potential-Outcomes Approach and Causal Effects of Interest**

The overall goal of our study was to estimate the causal effects of the initial nonattainment designations on Medicare outcomes in 2001. Importantly, the salient question was not “Did air pollution and health outcomes change during the time following the nonattainment designations?” but rather “Are air quality and health outcomes different after the nonattainment designations from what they would have been if the designations had not occurred?”

More formally, we defined the causal effect of interest as the comparison between two sets of potential outcomes: those that would occur if areas were designated as nonattainment for PM<sub>10</sub> in 1990–1995 and those that would occur if the nonattainment designations had not occurred. Note

that we decided to forgo the use of potential-outcomes notation in our analysis of this case study. However, we did use the more formal potential-outcomes notation in our analysis of Case Study 2 (see below). We considered the comparison of potential outcomes only for locations actually designated as nonattainment, that is, the estimand of interest was what is known in the causal inference literature as the average treatment effect on the treated (ATT), corresponding to the question “What were the causal effects of nonattainment designations for areas that were designated as such during 1990–1995?”

As a secondary objective, we also aimed to characterize the anticipated causal pathway whereby the nonattainment designations impact health as a consequence of reducing PM<sub>10</sub> in 1999–2001. In this context, PM<sub>10</sub> is an intermediate outcome that is expected to simultaneously be affected by the nonattainment designations and have bearing on the Medicare health outcomes; therefore, standard regression adjustment for ambient PM<sub>10</sub> in 1999–2001 cannot permit estimation of causal effects (Rosenbaum 1984). Accordingly, we used principal stratification (Frangakis and Rubin 2002) to quantify the extent to which causal effects of the designations on health outcomes were (1) associative with causal effects of the designations on PM<sub>10</sub> versus (2) dissociative with causal effects of the designations on PM<sub>10</sub>. As mentioned above, an associative effect that is larger than the dissociative effect indicates that the causal effect of the designations on Medicare health outcomes is greater in areas where pollution was causally reduced than in areas where there is little or no effect of the designations on pollution. Dissociative effects that are similar in size to associative effects indicate that the health impact of the intervention is similar regardless of whether the intervention decreased ambient pollution, which suggests the presence of other causal pathways through which the designations affected health without changing average ambient PM<sub>10</sub> during 1999–2001. It should be noted that causal reductions in pollution were defined on the basis of whether pollution was lower with the nonattainment designation than it would have been without the designation, not whether pollution decreased across time.

As outlined earlier in the Methods section, these types of causal questions can be framed as a hypothetical two-armed experiment with an “intervention” arm, corresponding to the observed allocation of nonattainment designations, and a “control” arm, corresponding to the hypothetical scenario with no nonattainment designations. Potential outcomes under the intervention condition are observed in nonattainment areas. Potential outcomes under the hypothetical control condition (with no attainment designations) are not observed in nonattainment areas. In order to characterize what would have happened in nonattainment areas had the nonattainment designations not occurred, we used

observed pollution and health outcomes in attainment areas that were not subject to SIP measures. Thus, the attainment areas can be construed as a control group for studying the effect of the nonattainment designations. The obvious threat to validity of the decision to estimate causal effects of the nonattainment designations by comparing outcomes with attainment areas is that the designations were decidedly not randomly assigned and thus attainment areas share important differences with nonattainment areas, which is evident from Table 1. Our use of data on attainment areas to learn what would have happened in nonattainment areas required careful confounding adjustment.

**Estimation of Propensity Scores for Confounding Adjustment**

The propensity score (Rosenbaum and Rubin 1983; Stuart 2010) is a nearly ubiquitous tool for adjusting for confounding in order to estimate causal effects with observational data that do not enjoy the benefits of randomization. The motivation for using the propensity score in our study was to be able to construct groups of attainment and nonattainment locations that were comparable with respect to the covariates listed in Table 1 so that the comparison of outcomes in attainment and nonattainment areas did not suffer from confounding on the basis of these factors. The key assumption we adopted for confounding

adjustment is that of strong ignorability that the covariates listed in Table 1 constitute (or are proxies for) all factors that could confound comparisons between attainment and nonattainment areas. This assumption, which amounts to the familiar “no unmeasured confounding” assumption, is discussed below.

We used a logistic regression for the probability of a nonattainment designation to estimate the propensity score, with predicted probabilities from this model representing the estimated propensity scores. Covariates included in this propensity score model are those listed in Table 1. If the propensity score model is adequately specified and the factors in Table 1 comprise all the relevant confounders (i.e., if there is no unmeasured confounding), then average outcomes in attainment areas represent what would have happened (on average) in nonattainment areas with similar values of the propensity score — as would be the case in a randomized study.

Figure 6 (left panel) depicts the distribution of estimated propensity scores for all 547 attainment and nonattainment locations. As expected, locations designated as nonattainment tend to have higher estimated propensity scores, but it should be noted that a wide range of estimated propensity scores are represented by both attainment and nonattainment areas. But locations with estimated propensity scores greater than 0.98 are exclusively nonattainment

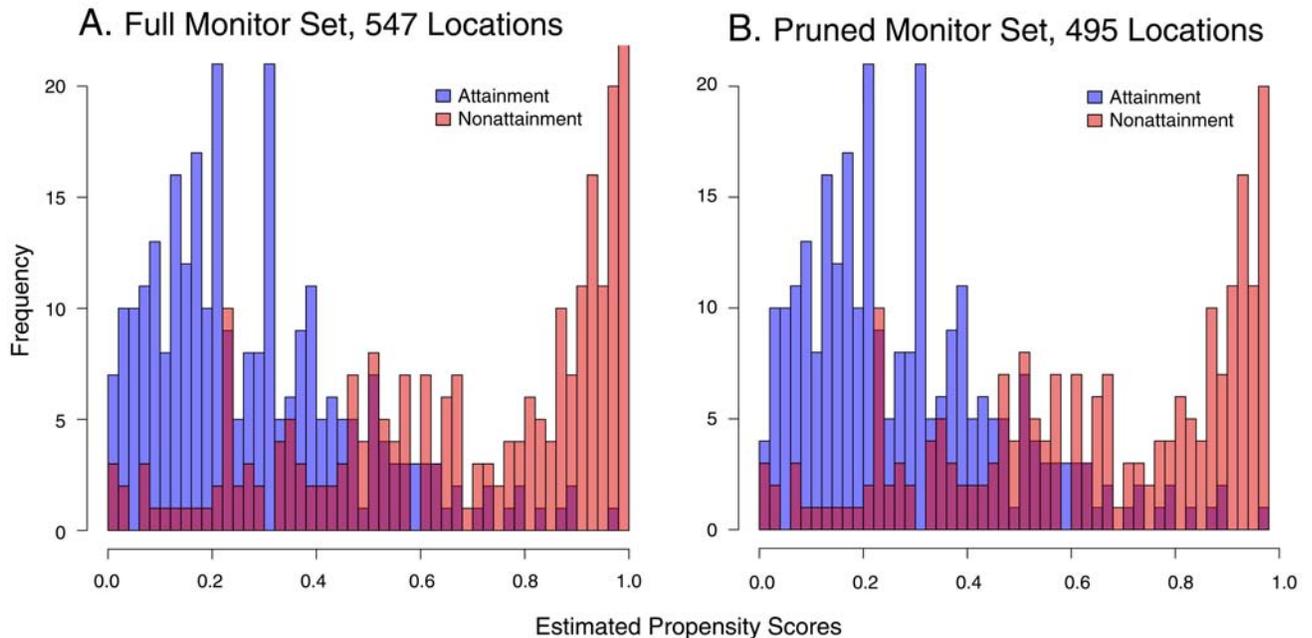


Figure 6. Histograms of estimated propensity scores for attainment and nonattainment areas before and after pruning observations with non-overlapping propensity score estimates. (Left) Full monitor set, 547 locations. (Right) Pruned monitor set, 495 locations.

locations. This phenomenon, where areas of the estimated propensity score distribution only have representation from either the treated or the control group, is sometimes referred to as a lack of overlap between propensity score distributions (Crump et al. 2009). Areas of the propensity score distribution that lack overlap are not appropriate for making causal inferences, and observations lying in those areas should be removed, or “pruned,” from the analysis data set to prevent model-based extrapolation beyond the range of observed data (Crump et al. 2009; Ho et al. 2007).

In this instance, 49 nonattainment locations had estimated propensity scores greater than 0.98 (i.e., these observations did not overlap with estimated propensity scores from attainment areas). The implication of this lack of overlap is that each of these 49 nonattainment areas exhibits a constellation of the covariates (see Table 1) that does not resemble that of any attainment area and thus provides no observed information from which to learn what would have happened in these areas had they not been designated. Put another way, these observations lack appropriate “control” observations. Analogously, three attainment areas had estimated propensity scores that did not overlap with those estimated in nonattainment areas. This lack of overlap is not surprising, as we would expect that, for example, there are areas of California’s Central Valley with population demographics and pollution levels that do not resemble those of any other part of the country. For estimation of causal effects that do not rely on model-based extrapolation of the confounding adjustment, we discarded (or pruned) the 52 observations without overlapping propensity scores (Ho et al. 2007; King and Zeng 2006). The pruning yielded an analysis sample that includes 495 monitoring locations, 219 of which are in nonattainment areas, with propensity score distributions depicted in the right panel of Figure 6. Figure 5 shows the locations of both the retained and pruned monitoring locations. This pruning reduced the study population to 3,555,934 Medicare beneficiaries. Strictly speaking, pruning such observations meant that estimates from our analysis could not technically be regarded as ATTs, as subsequent estimates only pertained to the subset of retained nonattainment locations. We assessed sensitivity to this pruning in an analysis that skips this pruning step and includes all 547 locations (the results of this analysis are in Appendix B, which is available on the HEI Web site).

After pruning the sample as described above, confounding adjustment was accomplished by grouping locations into five subclasses based on the quintiles of the estimated propensity score, because locations with similar values can be regarded as similar on the basis of all observed confounders. We classify the pruned analysis sample into five subgroups based on the quintiles of the

estimated propensity score, each subgroup containing attainment and nonattainment locations that have similar values of the propensity score (i.e., are comparable with regard to the factors in Table 1). Table 2 lists the number of attainment and nonattainment areas in each propensity score subclass. After adjustments for the propensity score subclass, any model for pollution and health outcomes can be used to estimate causal effects in a manner that is much less susceptible to observed confounding (Ho et al. 2007).

### Checking Covariate Balance

One essential benefit of using propensity scores in this analysis was that it was possible to check the extent to which grouping observations based on estimated propensity scores “worked,” in the sense that it ensured the construction of groups of attainment and nonattainment locations that were in fact comparable on the basis of the covariates listed in Table 1. If covariates were balanced between attainment and nonattainment areas within a propensity score subclass, the potential for these covariates to confound the analysis of causal effects was greatly reduced. One common metric for checking covariate balance is the standardized difference between attainment and nonattainment observations (Stuart 2010). This quantity can be calculated for each covariate as a way to summarize whether a covariate is in fact balanced between attainment and nonattainment areas, with values closer to zero indicating better average balance (and less susceptibility to confounding). We calculated standardized differences between attainment and nonattainment locations for each covariate before employing the propensity score (for all 547 monitoring locations in the original data set) (Figure 7). These standardized differences in the unadjusted sample can be regarded as a measure of the potential for bias in causal effect estimates caused by differences between attainment and nonattainment locations. We also calculated the standardized difference for each covariate within a propensity-score subclass as a measure of the covariate similarity

**Table 2.** Number of Attainment and Nonattainment Areas in Each of the Five Propensity Score Subclasses Used for Confounding Adjustment

Type of Area	Propensity Score Quintile				
	1st	2nd	3rd	4th	5th
Attainment	87	78	71	33	7
Nonattainment	12	21	28	66	92
Total	99	99	99	99	99

among attainment and nonattainment locations with similar values of the propensity score. The average of this value for each covariate across the five propensity score subclasses is shown in Figure 7. From the standardized differences plotted in Figure 7, we concluded that, despite the stark differences between attainment and nonattainment locations in the entire sample (the red line), the propensity score did an adequate job of balancing the covariates (listed in Table 1) between attainment and nonattainment locations, within propensity score subclass (the blue line).

### Models for Estimating Causal Effects

The preceding discussion of propensity scores did not involve Medicare health outcomes; nor did it pertain to any particular statistical model for actually estimating causal effects. Rather, we only formalized the causal effects of interest (ATTs), formulated the relevant “treatment” and “control” groups, and employed propensity scores to construct an analysis data set that would serve as the basis for estimating causal effects. After defining the five propensity score subclasses, we used parametric models for potential

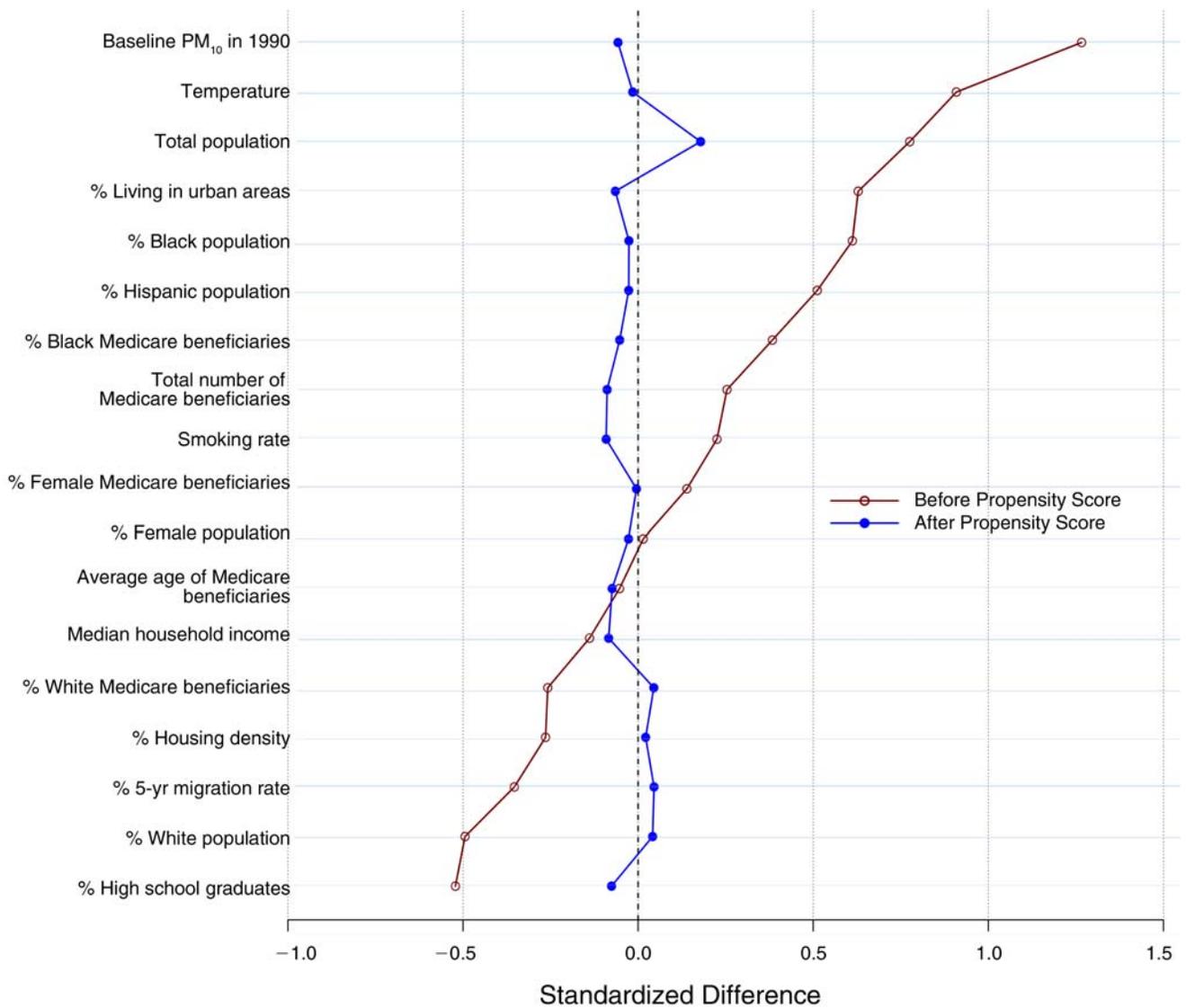


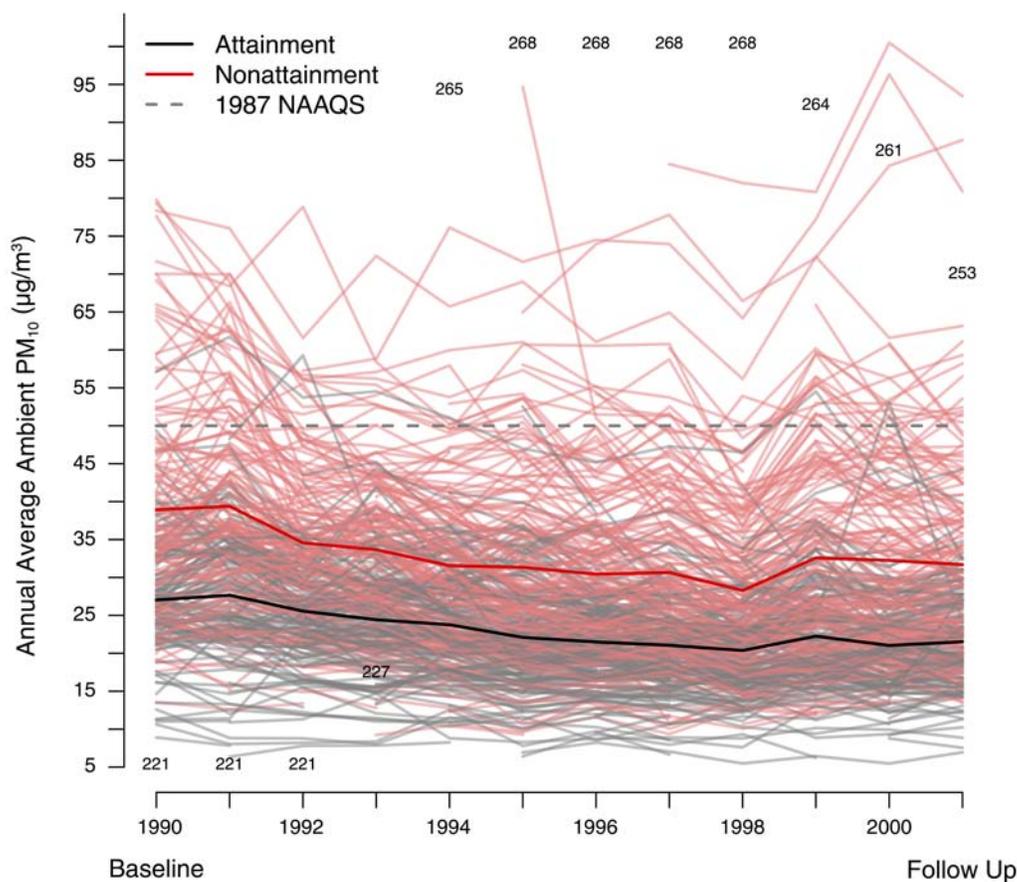
Figure 7. Description of covariate balance before (red line) and after (blue line) propensity score subclassification, as summarized by average standardized differences between nonattainment and attainment areas across each available covariate (from Table 1).

outcomes under attainment and nonattainment designations to predict potential outcomes that are not observed, namely, the potential pollution and health outcomes that would have occurred in nonattainment areas had the designations never occurred. Insofar as these predictions can be regarded as an accurate reflection of what would have happened in the absence of the designations, they could be used to estimate causal effects. Our analysis relied on two such models: (1) a spatial hierarchical regression model for (log-transformed) ambient  $PM_{10}$  concentrations during 1999–2001 and (2) log-linear Poisson regression models for each Medicare mortality and hospitalization outcome. All regression models were adjusted for propensity score subclasses and also for individual variables from Table 1 in order to adjust for any residual confounding not accommodated by the propensity score model and to improve efficiency. The spatial hierarchical model for pollution outcomes leveraged spatial correlation in ambient  $PM_{10}$  measures to further inform predictions of unobserved

potential outcomes. Details of these model specifications, the Markov chain Monte Carlo procedure used for estimation, and additional technical details can be found in work by Zigler and colleagues (2012) and in Appendix A.

### CASE STUDY 1: RESULTS

Figure 8 depicts the annual average ambient  $PM_{10}$  concentration (in  $\mu g/m^3$ ) for each of the 547 monitors for 1990–2001. As expected, annual ambient average  $PM_{10}$  concentrations in the early 1990s tended to be higher in nonattainment areas, but both attainment and nonattainment areas had annual ambient  $PM_{10}$  concentrations below the annual NAAQS during this time frame. Nonattainment locations with annual ambient  $PM_{10}$  concentrations below the annual standard may have violated the 24-hour NAAQS for  $PM_{10}$ , a point discussed below. It should be noted that ambient average  $PM_{10}$  decreases were similar in both attainment and nonattainment areas.



**Figure 8. Trends in annual average ambient  $PM_{10}$  from 1990 to 2001.** Thin lines represent individual monitoring locations, thick lines represent the average across all locations. The total number of nonattainment areas in each year is shown on graph. Note that the 1987 NAAQS is the annual  $PM_{10}$  standard, but many nonattainment areas may also have been in violation of the 24-hour  $PM_{10}$  standard.

### Unadjusted Comparisons

Analyses conducted using data on the entire sample of 547 monitoring locations indicated that between the baseline and follow-up time periods average ambient PM<sub>10</sub> decreased by 8.8 µg/m<sup>3</sup> in nonattainment areas (from 40.4 µg/m<sup>3</sup> in 1990 to 31.6 µg/m<sup>3</sup> in 1999–2001). In attainment areas the decrease was 5.4 µg/m<sup>3</sup> (from 27.0 µg/m<sup>3</sup> in 1990 to 21.6 µg/m<sup>3</sup> in 1999–2001). The *P* value comparing these changes from a two-sample *t*-test was *P* < 0.001. Among Medicare beneficiaries residing near one of the 547 monitors, the average rate of all-cause mortality (per 1000 person-years) in 2001 was similar in nonattainment and attainment areas (62.5 versus 62.6, *P* value from two-sample *t*-test *P* = 0.952). The average rate of CVD-related hospitalizations (per 1000 person-years) in 2001 was higher in nonattainment areas (92.1 versus 83.7, *P* value from two-sample *t*-test *P* < 0.001). Average rates of respiratory-related hospitalizations were similar in nonattainment areas and attainment areas (28.4 versus 28.4, *P* = 0.991). These unadjusted comparisons were likely confounded by differences between attainment and nonattainment areas.

### Average Causal Effects on Average Annual Ambient PM<sub>10</sub> in 1999–2001

Using the propensity score approach outlined above and confining interest to the ATT among the 219 nonattainment areas in the pruned sample, we estimated the causal effect of the nonattainment designations on average ambient PM<sub>10</sub> during 1999–2001 using the spatial hierarchical model outlined in Appendix A (section A.3), adjusted for the propensity score subclass and the variables in Table 1. The estimated causal effect of the nonattainment designations on average ambient PM<sub>10</sub> during 1999–2001 was  $-1.17$  µg/m<sup>3</sup> (95% posterior interval;  $-7.33, 4.00$ ). This decrease indicated that, among the 219 nonattainment areas, average ambient PM<sub>10</sub> during 1999–2001 was slightly lower than it would have been if the nonattainment designations had not occurred — that is, there is some evidence that the nonattainment designations had a causal effect on the 3-year average ambient PM<sub>10</sub> for the 1999–2001 period. However, this decrease cannot be considered statistically significantly different from zero. This result highlighted the likely possibility that decreases in this measure of PM<sub>10</sub> during this time frame (as evident from Figure 8) were likely caused in part by factors that affected both attainment and nonattainment areas. This point is discussed in more detail in the Conclusion and Discussion of this section.

### Average Causal Effects on Medicare Health Outcomes

For the Medicare health outcomes, we used the propensity score approach outlined above and the models outlined

by Zigler and colleagues (2012) and found in Appendix A (section A.3) to estimate ATTs among the 219 nonattainment locations in the pruned sample. Models used for predicting potential outcomes adjusted for propensity score subclass and the variables indicated in Table 1. Figure 9 summarizes posterior distributions of the average causal effects of the nonattainment designations on Medicare mortality (per 1000 beneficiaries) and hospitalization rates (per 1000 person-years) for CVD and respiratory issues. For all-cause mortality, the posterior mean ATT was  $-1.08$  deaths per 1000 Medicare beneficiaries (95% posterior interval;  $-3.27, 0.99$ ), suggesting that the nonattainment designations caused a decrease in mortality (i.e., that the average mortality rate in nonattainment areas was 1.08/1000 beneficiaries lower in 2001 than it would have been had these areas not been designated nonattainment). For CVD hospitalizations, the posterior mean ATT was 1.44 hospitalizations per 1000 person-years (95% posterior interval;  $-4.64, 7.16$ ), which did not indicate any causal effect on CVD-related hospitalizations among Medicare beneficiaries compared with what would have occurred without the designations. For respiratory hospitalizations, the posterior mean average causal effect of the nonattainment designations was  $-1.47$  (95% interval;  $-3.86, 0.70$ ) hospitalizations per 1000 person-years, indicating that the nonattainment designations causally reduced respiratory-related hospitalizations, compared with what would have occurred without the designations. All ATT estimates had 95% uncertainty intervals that included zero. The results of the analysis of CVD-related hospitalizations exhibited the most uncertainty.

### Associative and Dissociative Effects

To provide some insight into the existence of the anticipated causal pathway whereby the nonattainment designations might decrease PM<sub>10</sub>, which in turn might cause improvement in Medicare health outcomes, we employed the same models described above (conditional on the propensity score and covariates in Table 1) to estimate associative and dissociative effects (Table 3). Dissociative effects in this context are the causal effects of nonattainment designations on health outcomes among locations that were estimated to have experienced little or no causal effect on ambient PM<sub>10</sub> during 1999–2001. “Little or no” effect on ambient PM<sub>10</sub> was defined to include monitoring locations where the estimated causal effect on this measure of PM<sub>10</sub> was less than 5 µg/m<sup>3</sup>. Associative effects in this context are the causal effects of nonattainment designations on health outcomes among locations where ambient PM<sub>10</sub> during 1999–2001 was estimated to have decreased substantially as a result of the designation. A “substantial decrease” was defined to include monitoring locations where the estimated causal effect on this measure of PM<sub>10</sub> was a decrease

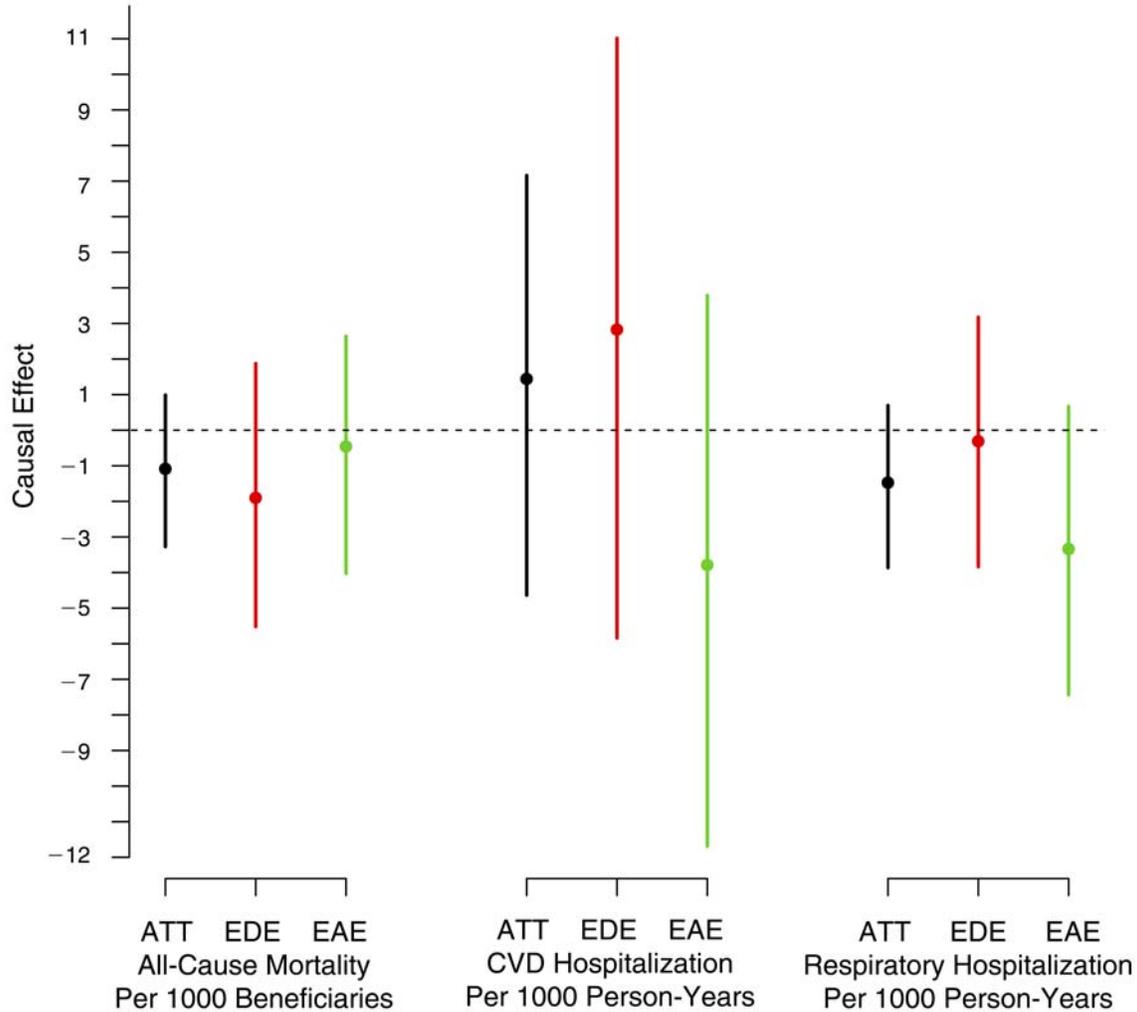


Figure 9. Posterior mean point estimates and 95% posterior intervals for the overall average effect on the treated (ATT, black), expected dissociative effect (EDE, red), and expected associative effect (EAE, green) of nonattainment designations on Medicare health outcomes from Case Study 1.

**Table 3.** Causal Effect Estimates for Overall, Associative, and Dissociative Effects in the Analysis of PM<sub>10</sub> Nonattainment Designations

Outcome	Overall Average Causal Effect			Average Dissociative Effect			Average Associative Effect		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
Ambient PM <sub>10</sub>	-1.17	-7.33	4	-1.9	-5.52	1.87	-0.46	-4.03	2.64
All-cause mortality	-1.08	-3.27	0.99	-1.9	-5.52	1.87	-0.46	-4.03	2.64
CVD hospitalization	1.44	-4.64	7.16	2.83	-5.84	11.01	-3.78	-11.69	3.79
Respiratory hospitalization	-1.47	-3.86	0.7	-0.31	-3.84	3.18	-3.34	-7.43	0.67

of at least  $5 \mu\text{g}/\text{m}^3$ . As a sensitivity analysis, average dissociative and associative effects were also estimated with the alternative value of  $1.2 \mu\text{g}/\text{m}^3$  (the approximate average effect on  $\text{PM}_{10}$ ) to define “little or no” and “substantial decrease” in  $\text{PM}_{10}$ .

For the mortality outcome, the posterior mean average dissociative effect was  $-1.90$  (95% interval;  $-5.52, 1.87$ ) deaths per 1000 beneficiaries, providing some evidence that mortality was reduced even in areas where  $\text{PM}_{10}$  during 1999–2001 was not causally affected. The posterior mean associative effect was  $-0.46$  (95% interval;  $-4.03, 2.64$ ), indicating no evidence of a causal effect on mortality in locations where  $\text{PM}_{10}$  is estimated to have been causally decreased by more than  $5 \mu\text{g}/\text{m}^3$ . The analogous estimates of dissociative effects and associative effects with the  $1.2 \mu\text{g}/\text{m}^3$  cutoff were  $-1.92$  (95% interval;  $-9.73, 6.69$ ) and  $-0.72$  (95% interval;  $-3.72, 1.92$ ), respectively. Note that the alternative definitions of dissociative and associative effects resulted in similar point estimates, but with wider uncertainty for the dissociative effect (which was calculated among fewer locations) and narrower uncertainty intervals for associative effect (which was calculated among more locations). This pattern of comparing associative and dissociative under the two different cutoff values persisted for all Medicare outcomes.

For the CVD-related hospitalization outcome, the posterior mean dissociative effect was  $2.83$  (95% interval;  $-5.84, 11.01$ ) hospitalizations per 1000 person-years, providing no evidence that CVD hospitalizations were causally affected in areas where  $\text{PM}_{10}$  during 1999–2001 was not causally decreased; the mean effect was actually suggesting an increase in CVD hospitalizations in these areas. The posterior mean associative effect was  $-3.78$  (95% interval;  $-11.69, 3.79$ ), which provided little evidence of a causal reduction in CVD hospitalization in locations where  $\text{PM}_{10}$  was estimated to have been causally decreased by more than  $5 \mu\text{g}/\text{m}^3$ . The analogous dissociative and associative effects with the  $1.2 \mu\text{g}/\text{m}^3$  cutoff were  $2.66$  (95% interval;  $-14.85, 21.03$ ) and  $-3.04$  (95% interval;  $-10.25, 3.39$ ), respectively. The uncertainty intervals were wide for both dissociative and associative effects for CVD-related hospitalizations.

For the respiratory-related hospitalization outcome, the posterior mean dissociative effect was  $-0.31$  (95% interval;  $-3.84, 3.18$ ) hospitalizations per 1000 person-years, providing no evidence that respiratory hospitalizations were causally affected in areas where  $\text{PM}_{10}$  during 1999–2001 was not causally decreased. In contrast, the posterior mean associative effect was  $-3.34$  (95% interval;  $-7.43, 0.67$ ), indicating a causal reduction in respiratory hospitalizations in locations where  $\text{PM}_{10}$  was estimated to

have been causally decreased by more than  $5 \mu\text{g}/\text{m}^3$ . Among the three outcomes, only the analysis of respiratory hospitalizations indicated an associative effect that was larger in magnitude than the dissociative effect, which was suggestive of the anticipated causal pathway; respiratory hospitalizations were not estimated to have been affected in areas where  $\text{PM}_{10}$  was not substantially affected, and these hospitalizations were estimated to have been causally reduced in areas where  $\text{PM}_{10}$  was causally reduced compared with what would have occurred without the nonattainment designations. The analogous dissociative and associative effects with the  $1.2 \mu\text{g}/\text{m}^3$  cutoff were  $-0.31$  (95% interval;  $-8.07, 7.04$ ) and  $-2.70$  (95% interval;  $-6.18, 0.56$ ), respectively.

### CONCLUSION AND DISCUSSION OF CASE STUDY 1

We employed the principles of the causal inference perspective (described above in the Methods section) to provide the first direct health-outcomes accountability assessment of one key feature of air pollution regulatory policy in the United States: the initial  $\text{PM}_{10}$  nonattainment designations that followed from the 1990 CAA amendments. Using a potential-outcomes perspective, we explicitly defined and estimated the causal effects of this specific set of regulatory decisions. Although ambient  $\text{PM}_{10}$  decreased in both attainment and nonattainment areas during the time frame of study, our results provided some evidence that 3-year average ambient  $\text{PM}_{10}$  concentration during 1999–2001 among areas designated as nonattainment in 1990–1995 was lower as a result of the nonattainment designations than it would have been if the designations had never occurred. Despite the modest effect on average ambient  $\text{PM}_{10}$  during 1999–2001, our results provided evidence that the nonattainment designations causally reduced mortality and respiratory-related hospitalizations among Medicare beneficiaries residing near a monitor located in a nonattainment area, as compared with what would have occurred if the nonattainment designations had not taken place.

The results from our investigation of the presumed causal pathway whereby nonattainment designations improved health outcomes through reducing ambient  $\text{PM}_{10}$  differed depending on the outcome of interest. The principal stratification analysis of the respiratory-related hospitalization outcome indicated an associative effect that was much larger in magnitude than the dissociative effect. The average dissociative effect near zero indicated that the nonattainment designations did not cause reductions in hospitalizations among areas where  $\text{PM}_{10}$  was not substantially causally affected. The average associative effect, which is different from zero, indicated a causal reduction in respiratory hospitalizations in areas where

PM<sub>10</sub> was estimated to have been causally reduced by more than 5 µg/m<sup>3</sup>. The result indicating that respiratory hospitalizations were causally reduced only when PM<sub>10</sub> was also causally reduced (i.e., a large associative effect relative to the dissociative effect) suggested the anticipated causal pathway. However, this principal stratification analysis could not conclusively indicate the improvement in health outcomes was attributable to the causal reduction in PM<sub>10</sub>, as would be the case in a formal mediation analysis (VanderWeele 2009). In particular, our analysis could not rule out the possibility that the correspondence between effects on health and effects on pollution was driven by a factor other than the nonattainment designations or discern whether health effects are the result of an alternative causal pathway present within the areas where ambient PM<sub>10</sub> during 1999–2001 was causally reduced. (We present a formal causal mediation analysis in the section describing Case Study 2.) Nonetheless, we argue that the finding that the causal effect of PM<sub>10</sub> nonattainment designations on respiratory hospitalization outcomes was most pronounced in areas exhibiting causal reductions in ambient PM<sub>10</sub> is useful for informing future policies.

The principal stratification analysis of the CVD-related hospitalization outcome showed a similar pattern, with causal effects on this outcome most strongly pronounced in areas where the nonattainment designations decreased PM<sub>10</sub>; however, all estimates for this outcome were subject to substantial uncertainty. The principal stratification analysis of the mortality outcome provided no evidence that dissociative effects were different from associative effects, which suggested that any evident causal effects of nonattainment designations on mortality were likely the result of causal pathways other than the impact on 3-year average ambient PM<sub>10</sub> in 1999–2001. Examples of other important causal pathways include simultaneous or synergistic impacts of the intervention on other pollutants, the initiation of economic consequences that affect health outcomes in the long term, or even other measures of PM<sub>10</sub>, possibly in a different time frame (HEI Accountability Working Group 2003).

A key limitation of our analysis is the fact that we estimated the effect of the nonattainment designation by regarding all monitoring locations in a nonattainment area as “treated.” However, nonattainment designations resulted in a wide variety of specific actions to control air quality on state and local levels, and sometimes resulted in no action at all. A nonattainment designation may not have prompted any action in a SIP when, for example, regional control strategies were expected to reduce PM<sub>10</sub> to achieve the NAAQS regardless of the nonattainment designations. Monitoring locations that were in nonattainment counties but

didn’t themselves indicate elevated PM<sub>10</sub> were likely not the target of localized actions to control PM<sub>10</sub> (Auffhammer et al. 2009). For example, we found evidence that the causal effect of a designation on average ambient PM<sub>10</sub> was most pronounced in areas that had the highest annual concentrations in 1990. For example, the effect calculated only among the 35 monitors in nonattainment areas with baseline annual concentration of more than 40 µg/m<sup>3</sup> was –5.8 µg/m<sup>3</sup> (95% posterior interval; –22.8, 7.5). Furthermore, the analysis did not distinguish between locations that were designated by means of the annual or 24-hour standard, which may have also determined any course of action in a SIP.

One consequence of framing our approach in a formal potential-outcomes framework was the need to precisely define a specific action or decision that corresponded to the “intervention” of interest, here the nonattainment designations. A resulting limitation is that the analysis estimated causal effects that were averaged over various different types of monitoring locations in nonattainment areas subject to a diversity of actions (or inaction). In this sense, estimating the causal effects of nonattainment designations was akin to an “intention to treat” analysis in clinical studies that consider causal effects of assignment to an intervention, as opposed to actual receipt of that intervention. Available data on specific actions taken on local scales (e.g., measures in a SIP) could facilitate a causal analysis of these actions, but their use for informing future policy could be limited because specific control actions are often highly specialized to local circumstances (e.g., spraying wind-blown dust in Central California) and may not be replicable or relevant in other areas or at future time points. In focusing on the causal effects of a set of federal-level regulatory decisions, we exchanged some level of detail with regard to actual control measures for precision in defining the intervention and evidence of the effectiveness of a regulatory process that can (and will) be replicated in the future. Estimating the effects of the designations, and thus averaging over a variety of possible control actions, may explain modest point estimates (e.g., in the estimate of the effect on annual average PM<sub>10</sub> in 1999–2001) and also the high level of estimation uncertainty.

A related limitation is the likely possibility that the reductions in ambient PM<sub>10</sub> observed between the early 1990s and early 2000s were attributable to other control measures that exist outside of the paradigm of nonattainment designations. Many areas likely started to take action to improve air quality before 1990 in anticipation of the impending designations, and any reductions in PM<sub>10</sub> during this time would not be reflected in our analysis, which relied on data dating back only as far as 1990. Furthermore, many of the initial PM<sub>10</sub> nonattainment areas had

been taking action towards attainment of the standard for total suspended particles, which predated the PM<sub>10</sub> standard, and such actions likely had an effect on PM<sub>10</sub>. Control measures put in place before 1990 may have contributed to a muting of the estimates of the effects of the nonattainment designations in comparison with expectations.

As with virtually all air pollution epidemiology, one significant challenge to causal inference is the prospect of confounding. In the accountability context considered here, confounders were factors that differed between attainment and nonattainment areas that also bore some relationship with pollution or health outcomes. In studies of exposure–response relationships the confounders are generally considered to be factors jointly associated with pollution and health outcomes. Observed confounding in our context was particularly pronounced, as nonattainment areas were designated precisely because they exhibited (or contributed to) poor air quality, which was associated with a multitude of factors that differentiated attainment and nonattainment areas. Our propensity score strategy was able to group attainment and nonattainment locations that were similar on the basis of baseline pollution levels, characteristics of the Medicare beneficiaries residing within 6 miles of an air quality monitor, and numerous features of the general nearby population, thus minimizing the chance of confounding with regard to these factors. Importantly, we discarded monitor locations in nonattainment areas that were not similar to any attainment area on the basis of observed confounders (and vice versa). Including these locations in our analysis would have estimated causal effects that necessarily relied on model extrapolation beyond the information contained in the observed data, while removing them restricted our conclusions to only a subset of nonattainment areas. By way of comparison, we did a sensitivity analysis using the same methods, but without the pruning of observations (Appendix B, available on the Web), an analysis expected to be particularly susceptible to confounding because of the extrapolation of inference to nonattainment areas with no comparable attainment area in the data. Compared with the main analysis, the analysis in Appendix B estimated a more pronounced causal reduction in ambient PM<sub>10</sub>, larger causal reductions in all-cause Medicare mortality rates, point estimates indicating larger causal increases in CVD-related hospitalizations (with very wide uncertainty), and effects on respiratory-related hospitalizations that were very similar to those in the main analysis. Although our strategy is specifically designed to mitigate bias attributed to measured confounders, the prospect of unmeasured confounding remained a threat to the validity of our results. If unmeasured factors related to pollution and/or health outcomes exist that, even after adjustment for all observed factors listed in Table 1, still differ between

attainment and nonattainment areas, then our results are subject to unmeasured confounding.

Finally, our analysis did not explicitly account for the regional nature of air quality control. Because issues such as regional pollution transport, actions undertaken in SIPs in particular areas may have impacts that spread across to other areas. More broadly, regional transport of air pollution often leads to regional control strategies that simultaneously affect both attainment and nonattainment locations. For example, California adopted statewide fuel and other restrictions that reduced PM<sub>10</sub> and precursor emissions from mobile sources during this same time frame. The likelihood that actions in nonattainment areas could have effects that spill over into nearby attainment areas would dilute the causal effects we aimed to estimate. Pollution and health outcomes in attainment areas likely improved as well, and the present analysis did not account of this improvement. Thus, pollution and health outcomes under a setting where nonattainment designations had never occurred may have actually been worse than was indicated by the “control group” of our analysis when that group may have experienced benefits of the designations in other areas. This phenomenon is known as interference in the causal inference literature. Our work in Zigler and colleagues (2012) outlined an assumption about interference under which our approach would be robust to the effects of regional pollution transport. The assumption relied on the feature that nonattainment designations implicitly considered regional transport to some extent in that a particular area’s designation could be based on contributions to air quality in other areas, although this was not expected to be as important for actions controlling PM<sub>10</sub> as those targeting other pollutants, such as PM<sub>2.5</sub> or O<sub>3</sub>.

Our analysis made use of a vast data resource that links together information on regulatory actions, ambient air quality, population characteristics, and health information on the entire U.S. Medicare enrollment population. Accountability assessment of large-scale regulatory interventions relies on the large-scale availability of health data such as the administrative hospital claims available on millions of individuals. Our analysis focused on the susceptible population of the elderly in the United States that lived near a PM<sub>10</sub> monitor in the EPA monitoring network. Alternative pollution measurement techniques (e.g., satellite measurements or spatial extrapolation) could expand such an analysis to consider individuals not residing close to a monitor location, and other administrative data sources (e.g., hospital emergency department records or electronic health records) could be used to focus on other populations.

Nonattainment designations are one key mechanism for air quality management in the United States and represent a key step in the achievement of the NAAQS. Although our

analysis entailed important limitations, it provided evidence of the effectiveness of this integral feature of air quality management in the United States and represents a distinct perspective that should be interpreted in conjunction with — not instead of — the large body of epidemiological research motivating the setting and implementation of NAAQS.

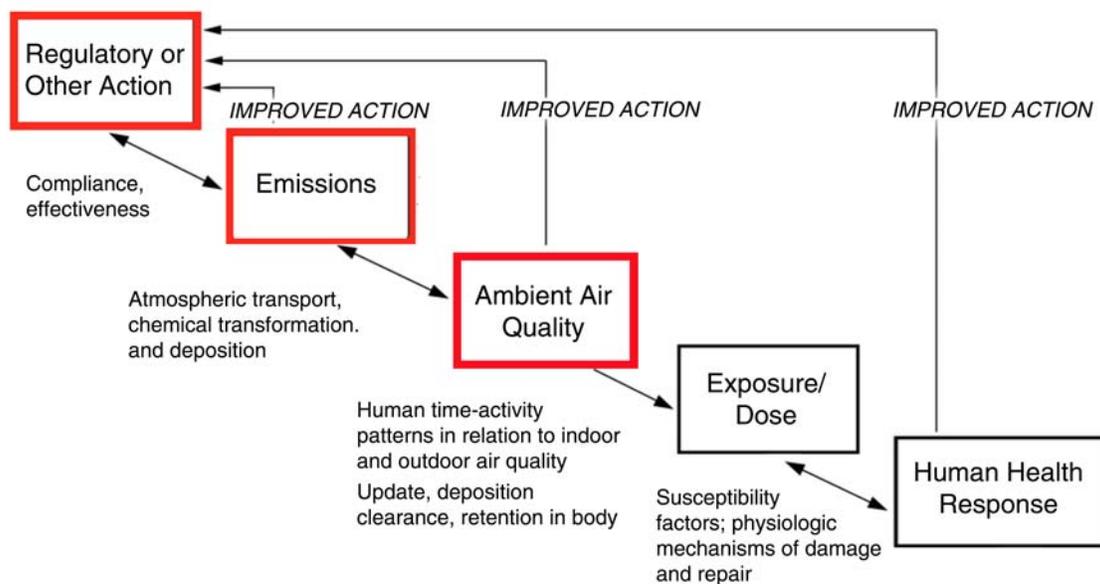
### Additional Work in Progress

In addition to the case study presented above, we have engaged in a variety of related research endeavors designed to improve direct-accountability assessment of the PM<sub>10</sub> nonattainment designations. In Cefalu and Zigler (2015), we proposed new methodology that generalizes the approach of omitting observations that lack propensity score overlap to allow for a stochastic filtering that ultimately weights causal estimates according to posterior evidence that each observation has a comparable observation in the opposite treatment group. We are working to deploy this methodology to analyze the causal effects of nonattainment designations over time. We are also working to corroborate the results of the analysis presented here with newly developed methods for Bayesian nonparametric principal stratification and causal mediation analysis in a multipollutant context.

### CASE STUDY 2: ACCOUNTABILITY ASSESSMENT OF POWER-PLANT EMISSIONS CONTROLS

In the second case study of this report, we investigated the causal impacts of SO<sub>2</sub> scrubbers in coal-fired power plants on multiple emissions and on ambient PM<sub>2.5</sub>. Three links in the chain of accountability are considered in this case study: regulatory action, emissions, and ambient air quality (Figure 10). The primary focus of this investigation was to illustrate the ideas outlined above in the section describing our causal pathways analyses as well as to demonstrate our newly developed methods for principal stratification and causal mediation analysis in a multipollutant setting. Accordingly, the discussion of Case Study 2 entails significantly more technical methodological detail than the previous sections of this report. The analysis of this case study should be viewed as proof of concept, in the sense that it deploys sophisticated new statistical methods to a problem that is relatively well understood and non-controversial — that of characterizing how scrubbers on coal-fired power plants reduce ambient PM<sub>2.5</sub>.

Various CAA regulations have required or otherwise resulted in the installation of SO<sub>2</sub> scrubbers in new or existing EGUs prior to the 1990 CAA amendments. These regulations include New Source Performance standards,



**Figure 10. Links in the chain of accountability considered in Case Study 2.** In our analysis Regulatory or Other Action refers to the presence of SO<sub>2</sub> scrubbers in 2005, Emissions to the annual emissions of SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub> in 2005, and Ambient Air Quality to the annual average ambient PM<sub>2.5</sub> concentration in 2005 among monitoring locations located within 150 km of a power plant. (Adapted from Figure 1, Health Effects Institute 2010.)

Prevention of Significant Deterioration requirements for major sources, and implementation of the national ambient standards for SO<sub>2</sub>. The 1990 amendments to the CAA added the ARP, which required major emissions reductions of SO<sub>2</sub> (and other emissions) from American power plants. One goal of this program was to reduce total SO<sub>2</sub> emissions by 10 million tons relative to 1980 levels of 29.5 million tons per year. This reduction was to be achieved mostly through cutting emissions from EGUs, a process enacted in two phases. Phase I (1995–1999) required 263 of the EGUs with the highest emissions to significantly reduce their emissions. Phase II, which began in 2000, established a target SO<sub>2</sub> emissions cap of 8.95 billion tons per year on about 3,200 EGUs, to cut power-sector emissions nearly in half from 1980 levels. Significantly, in implementing this market-based allowance trading program in both phases of the ARP, the CAA amendments required installation of continuous emissions monitors for SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub>. The emissions data obtained as a result have been described earlier in the section on regulatory data. Sources were permitted to choose whether and how to reduce SO<sub>2</sub> emissions as long as they met their progressive allowance requirements.

Impacts of the ARP have been evaluated extensively, and the program is generally lauded as a success story because marked national decreases in SO<sub>2</sub> and NO<sub>x</sub> were achieved at a relatively low cost. Despite a 25% increase in electricity production over the first 14 years of the program, SO<sub>2</sub> emissions decreased by 36% (U.S. EPA 2011). The program met its long-term goal of reducing EGU annual SO<sub>2</sub> emissions to 8.95 tons by 2007, with emissions decreasing further through at least 2010, in part because additional regional reduction requirements were promulgated by U.S. EPA for much of the Eastern United States (Schmalensee and Stavins 2012). Estimates of the annualized human health benefits of the entire ARP have ranged from \$50 billion to \$100 billion (Banzhaf et al. 2006; Burtraw 1999; Burtraw et al. 1998; Chestnut and Mills 2005; Schmalensee and Stavins 2012). Recent analyses have used air quality model simulations to provide more targeted estimates of the health benefits attributable to emissions reductions from EGUs (Buonocore et al. 2014; Levy et al. 2007, 2009). Whether attempting to quantify the health impacts of the ARP as a whole or provide analyses of emissions reductions from specific EGUs, the existing evidence of the health benefits relies heavily on presumed relationships between power-plant emissions, ambient PM<sub>2.5</sub>, and human health.

Although power plants had latitude under the ARP to elect a variety of strategies to reduce emissions, such as changes in combustion technology or shifts in fuel composition, one key strategy was the installation of scrubbers on

their EGUs to reduce SO<sub>2</sub> emissions. In this report, the term scrubber covers these SO<sub>2</sub> emissions control technologies: dry lime flue-gas desulfurization scrubbers, dry-sorbent injection scrubbers, dual-alkali scrubbers, magnesium-oxide scrubbers, sodium-based scrubbers, wet lime flue-gas desulfurization, wet limestone scrubbers, fluidized bed limestone injection, and technologies listed in the AMPD as “other” SO<sub>2</sub> control strategies. Although the ability of a scrubber to reduce ambient PM<sub>2.5</sub> through reducing SO<sub>2</sub> is largely regarded as known, the nature of these relationships has never been quantified empirically amid the realities of actual regulatory implementation, where pollution controls may impact a variety of factors that are also related to the formation of PM<sub>2.5</sub>.

The goal of this case study was to assess whether our newly developed methods would produce estimates that were consistent with the relatively well-understood effects on the ambient concentration of PM<sub>2.5</sub> of installing a scrubber on a coal-fired power plant. In particular, we sought to quantify and compare the contribution of the presumptive causal pathway in which a scrubber reduces SO<sub>2</sub> emissions, which in turn reduces ambient PM<sub>2.5</sub>, to the contribution of other causal pathways attributed to concurrent reductions (or co-benefits) in other emissions or other factors. Thus, the question was formally framed as one of mediation analysis: To what extent is the causal effect of a scrubber (the “treatment”) on ambient PM<sub>2.5</sub> (the “outcome”) mediated through reduced emissions of SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub> (the “mediators”)?

To answer these questions, we used the data sources described above in the Methods section to provide a more refined direct-accountability assessment of the extent to which a particular emissions-control action reduced emissions and caused improvements in ambient air quality. Specifically, we evaluated the extent to which an SO<sub>2</sub> scrubber on a coal-fired power plant (1) causally affected emissions of SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub>, (2) causally affected ambient PM<sub>2.5</sub>, and (3) caused effects on ambient PM<sub>2.5</sub> in a manner that was mediated through reducing SO<sub>2</sub>, NO<sub>x</sub>, and/or CO<sub>2</sub>. We focused in particular on the question of mediated effects to provide rigorous statistical evidence of the presumed relationships between actions, emissions, and ambient pollution that form the basis of a great deal of existing health benefits analyses. To this end, we used our newly developed Bayesian nonparametric statistical methods, which drew on two frameworks for estimating causal effects in the presence of intermediate mediating variables: (1) principal stratification (Frangakis and Rubin 2002) and (2) direct and indirect effects, or so-called causal mediation analysis (Pearl 2001; Robins and Greenland 1992; VanderWeele and Vansteelandt 2014). Both frameworks required the

development of new statistical methods to accommodate the multipollutant nature of the problem. The expectations a priori — based on existing knowledge of relationships among scrubbers, emissions, and ambient PM<sub>2.5</sub> — were that scrubbers lead to a pronounced causal reduction in SO<sub>2</sub> emissions but have little or no impact on emissions of NO<sub>x</sub> and CO<sub>2</sub>, and that any effect of scrubbers on ambient PM<sub>2.5</sub> is primarily mediated through reductions in SO<sub>2</sub> emissions.

### Linked Data Sources

Using the tools described above in the Methods section, we assembled a national database to use in the investigation. We obtained annual emissions data from the year 2005 from continuous emissions monitors on 258 coal-fired power plants. We also obtained information on characteristics of the plants, including various NO<sub>x</sub> emissions controls that may or may not have operated simultaneously with

scrubbers, the annual average heat input (in 2004), the percent operating capacity (calculated as heat input divided by plant total capacity), the participating phase of the ARP (I or II), operating time in 2004, and the coal sulfur content in 2004 (Table 4).

We took the statistical unit of analysis to be individual power plants, which consist of a collection of at least one EGU. Ambient monitoring data was aggregated to the level of the power plant. To do this, we linked each power plant to all ambient PM<sub>2.5</sub> monitors located within a 150-km radius of the plant. Monitors located within 150 km of more than one plant were linked only to the closest plant. Figure 11 depicts the locations of the power plants and linked ambient PM<sub>2.5</sub> monitors. For each plant, we calculated the average annual ambient PM<sub>2.5</sub> concentration in 2005 among all PM<sub>2.5</sub> monitors linked to that plant, as well as average temperature and barometric pressure measured at the monitoring locations during 2004. This linkage

**Table 4.** Summary Statistics for Covariates and Outcomes Available for the Analysis of SO<sub>2</sub> Scrubbers

	Power Plants With Scrubbers ( <i>n</i> = 63)		Power Plants Without Scrubbers ( <i>n</i> = 195)	
	Mean	SD	Mean	SD
<b>Monitor Data<sup>a</sup></b>				
Average ambient PM <sub>2.5</sub> (µg/m <sup>3</sup> ) (2005)	11.60	4.00	13.20	2.60
Average temperature (Celsius) (2004)	12.90	4.40	13.10	3.90
Average barometric pressure (mm Hg) (2004)	718.40	52.00	743.30	22.80
<b>Power-Plant-Level Emissions (tons)</b>				
Total SO <sub>2</sub> emissions (2005)	1,293.50	1,801.20	2,165.50	2,545.90
Total NO <sub>x</sub> emissions (2005)	956.20	797.70	582.50	552.30
Total CO <sub>2</sub> emissions (2005)	568,656.00	464,987.50	372,380.70	369,165.50
<b>Unit Level Data</b>				
Number of NO <sub>x</sub> controls (Jan. 2004)	1.10	0.60	1.00	0.60
Selective catalytic or selective noncatalytic reduction (Jan. 2004) <sup>b</sup>	0.30	0.40	0.20	0.40
Operation relative to capacity (MMBTU input/maximum MMBTU capacity) (2004) <sup>c</sup>	20.60	11.10	17.70	10.10
Average heat input (MMBTU) (2004)	5,534,482.00	4,535,298.30	3,627,818.10	3,595,413.60
Phase II indicator (2004)	0.80	0.40	0.80	0.40
Total operating time (hr/yr) (2004) <sup>b</sup>	7,688.00	688.60	7,354.00	1,065.50
Sulfur content in coal (% by weight) (2004) <sup>b</sup>	1.40	1.00	0.80	0.60

<sup>a</sup> Calculated among monitors within 150 km radius of each power plant.

<sup>b</sup> These variables are weighted averages within a power plant, with weights according to each EGU's annual heat input.

<sup>c</sup> Measure is monthly percentage of operating capacity, summed over all months and over all EGUs within a power plant (values > 1.00).

strategy is shown in Figure E.1(a) of Appendix E, which is available on the Web. Constructing the data set in this way entailed many limitations, mostly because of the realities of regional pollution transport. The causal quantities being estimated are clarified in the following section and their limitations are discussed later in the Conclusion and Discussion section.

**Defining the Intervention: SO<sub>2</sub> Scrubber Installation**

Since a power-generating facility can consist of multiple EGUs, each of which may or may not be equipped with a scrubber, we must define the intervention at the level of the facility. To do this, we regarded a facility as “treated” with a scrubber if at least 10% of the total heat input for that facility can be attributed to EGUs within that facility

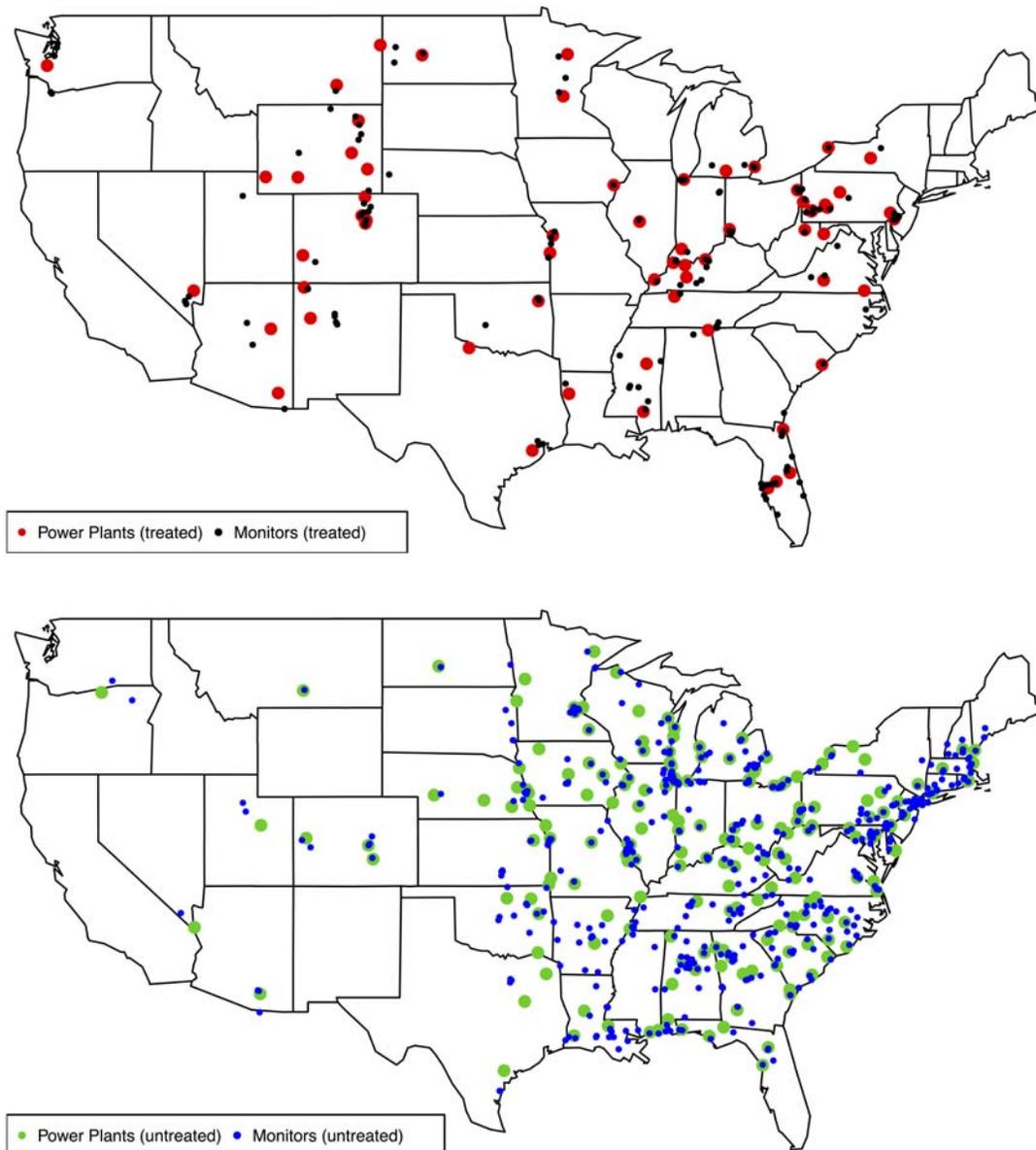


Figure 11. Locations of power plants and linked ambient PM<sub>2.5</sub> monitors for the analysis of SO<sub>2</sub> scrubber effects on emissions and ambient PM<sub>2.5</sub>.

that had scrubbers installed as of January 2005. Figure C.1, in Appendix C (available on the Web), shows the distribution of percentage of heat input from an EGU with a scrubber across all 258 facilities. The vast majority of facilities had nearly all or nearly none of their heat input attributed to EGUs with scrubbers. This distribution suggested that exactly how to define facilities as having been “treated” with a scrubbers was relatively unimportant. The “control” condition used for comparison was the setting where no scrubbers were installed. Thus, causal effects in this case study related to comparisons between emissions and ambient  $PM_{2.5}$  that would be potentially observed if a particular facility did or did not adopt scrubbers to control  $SO_2$  emissions. The “intervention group” of the study consisted of the 63 power plants that had scrubbers in January 2005. The “control group” consisted of the 195 plants that did not have scrubbers. Note that three power plants were excluded from this analysis because scrubbers had been installed during 2005.

### Defining Potential Outcomes for Principal Stratification and Causal Mediation Analysis

The primary objective of this case study was to characterize the extent to which installing a scrubber impacts ambient  $PM_{2.5}$ , “through” altering emissions of  $SO_2$ ,  $NO_x$ , and  $CO_2$ . We refined the previous descriptions of potential outcomes, principal stratification, and direct and indirect effects to accommodate multipollutant accountability settings.

We formulated the approach with explicit potential-outcomes notation (Rubin 1978) as follows. Consider a single power plant and let  $Z \in \{0,1\}$  denote whether the power plant had scrubber(s) installed in January 2005, with  $Z = 1$  denoting the presence of a scrubber. Let  $\{M_k(z); k = 1, \dots, K\}$  denote the potential emissions of  $K$  pollutants that would occur if the power plant were to have scrubber status  $Z = z$ , for  $z = 0,1$ . Henceforth, we fix  $K = 3$  so that  $M_k(z), k = 1,2,3$  denotes the potential emissions of  $SO_2$ ,  $NO_x$ , and  $CO_2$ , respectively. The causal effect of the scrubber on emission  $k$  can then be defined as a comparison between  $M_k(1)$  and  $M_k(0)$  comparing emissions that would be observed under the “treatment” and “control” conditions. Let  $\mathbf{M}(z_1, z_2, z_3) \equiv \{M_1(z_1), M_2(z_2), M_3(z_3)\}$  denote potential emissions under a set of three kinds of scrubber status  $\{z_1, z_2, z_3\}$ . For example,  $\mathbf{M}(1,0,0)$  would represent the potential  $SO_2$  emissions under installation of a scrubber and the potential  $NO_x$  and  $CO_2$  emissions that would be observed if the scrubber had not been installed.

We similarly defined potential  $PM_{2.5}$  outcomes, but extended the notation to define potential concentrations under different potential values of scrubber status,  $Z$ , and

different possible values of emissions,  $\mathbf{M}(z_1, z_2, z_3)$ . Thus, in full generality, each power plant has a set of  $2^4 = 16$  potential outcomes for  $PM_{2.5}$ ,  $Y(z; \mathbf{M}(z_1, z_2, z_3))$ , which denotes potential values of  $PM_{2.5}$  that would be observed under intervention  $Z = z$  with pollutant emissions set at values under interventions  $z_1, z_2, z_3$ . Defining all 16 potential  $PM_{2.5}$  concentrations implies that each emission could, at least in theory, be intervened upon independently of  $PM_{2.5}$  and the other emissions. Thus, it is worth noting that values of  $Y(z; \mathbf{M}(z_1, z_2, z_3))$ , can be categorized into two groups: those that are observable and those that are unobservable. For  $z = z_1 = z_2 = z_3$ , potential outcomes are observable from the data, that is, any power plant with a scrubber will have  $Y(1; \mathbf{M}(1,1,1))$  observed, and any power plant without a scrubber will have  $Y(0; \mathbf{M}(0,0,0))$  observed. We refer to these as *observable* potential outcomes. In contrast, potential outcomes defined under any other values of the vector  $(z, z_1, z_2, z_3)$  represent potential outcomes where a power plant is simultaneously subjected to different interventions, and can never be observed in practice. For example,  $Y(1; \mathbf{M}(0,0,1))$  represents the potential ambient  $PM_{2.5}$  concentration near a plant under the hypothetical scenario in which that plant installs a scrubber ( $z = 1$ ), but emissions of  $SO_2$  and  $NO_x$  are set to what they would be without the scrubber ( $z_1 = z_2 = 0$ ) and emissions of  $CO_2$  are set to what they would be with the scrubber ( $z_3 = 1$ ). We refer to these potential outcomes as *unobservable* (or a priori counterfactual), as they are never observed for any power plant. Estimating causal effects relying on such unobservable potential outcomes will rely on unverifiable assumptions that relate each of these unobservable quantities to observed relationships in the data.

Note that the total effect (TE) of scrubber installation on ambient  $PM_{2.5}$  can be defined as the comparison between the observable potential outcomes  $Y(1; \mathbf{M}(1,1,1))$  and  $Y(0; \mathbf{M}(0,0,0))$ . Various other causal effects related to causal pathways will be defined based on comparisons between different combinations of the above potential outcomes.

Principal stratification defines causal effects based only on the observable potential outcomes  $Y(1; \mathbf{M}(1,1,1))$  and  $Y(0; \mathbf{M}(0,0,0))$ . Associative effects represent causal effects of a scrubber on ambient  $PM_{2.5}$  among power plants where emissions are causally affected by the scrubber. Dissociative effects represent causal effects of a scrubber on ambient  $PM_{2.5}$  among power plants where emissions are not meaningfully affected by the scrubber. In the presence of multiple pollutants, associative and dissociative effects can be defined as functions of changes in each of the  $K = 3$  emissions. Following the development in Zigler and

colleagues (2012), we focused discussion on average (or expected) associative and dissociative effects defined as:

$$\text{EDE}_{\mathcal{K}} = E \left[ \begin{array}{l} Y(1; \mathbf{M}(1,1,1)) - Y(0; \mathbf{M}(0,0,0)) \mid \mid \\ (\mathbf{M}(0,0,0) - \mathbf{M}(1,1,1)) \mid_{\mathcal{K}} < \mathbf{C}_{\mathcal{K}}^D \end{array} \right], \quad (1)$$

$$\text{EAE}_{\mathcal{K}} = E \left[ \begin{array}{l} Y(1; \mathbf{M}(1,1,1)) - Y(0; \mathbf{M}(0,0,0)) \mid \mid \\ (\mathbf{M}(0,0,0) - \mathbf{M}(1,1,1)) \mid_{\mathcal{K}} > \mathbf{C}_{\mathcal{K}}^A \end{array} \right], \quad (2)$$

where  $\mid (\mathbf{M}(1,1,1) - \mathbf{M}(0,0,0)) \mid_{\mathcal{K}}$  denotes a vector of absolute differences between potential emissions of the subset of pollutants in the set  $\mathcal{K}$ , with  $>$  and  $<$  representing element-wise comparisons between vectors of mediators. For example,  $\mathcal{K} = \{1,2\}$  would be used to define associative and dissociative effects based only on causal effects on emissions of  $\text{SO}_2$  and  $\text{NO}_x$ , without regard to the effect on  $\text{CO}_2$ . Here,  $\mathbf{C}_{\mathcal{K}}^A$  denotes a vector of thresholds beyond which a change in each pollutant emission in  $\mathcal{K}$  is considered meaningful, whereas  $\mathbf{C}_{\mathcal{K}}^D$  is a vector of thresholds below which changes in these pollutant emissions are considered not meaningful. For example, with  $\mathcal{K} = \{1,2,3\}$  and  $\mathbf{C}_{\mathcal{K}}^A = \{1,2,3\} \equiv \{C_1^A, C_2^A, C_3^A\}$ , EAE could estimate the average causal effect on  $\text{PM}_{2.5}$  among power plants for which scrubber installation causally affected emissions of  $\text{SO}_2$ ,  $\text{NO}_x$ , and  $\text{CO}_2$  in excess of  $C_1^A$ ,  $C_2^A$ ,  $C_3^A$ , respectively. Estimates of EAE and EDE are useful summary measures of causal effects on  $\text{PM}_{2.5}$ , on average, when emissions change or do not change, but the relationship between causal effects on  $\text{PM}_{2.5}$  and causal effects on emissions can vary as an entire surface describing effects on  $\text{PM}_{2.5}$  for any particular values of  $\mathbf{M}(0)$  and  $\mathbf{M}(1)$ . In addition to estimating EDE and EAE for different  $\mathcal{K}$  as defined above, we also estimated entire surfaces of, for example, how the causal effect on  $\text{PM}_{2.5}$  varied as a function of the causal effect on each emission.

Causal mediation analysis relies on the definition of natural direct effects (NDE) and natural indirect effects (NIE) (Pearl 2001; Robins and Greenland 1992), which are defined based on potential outcomes that were described above as unobservable. NDE in this context were defined as causal effects of scrubber installation on  $\text{PM}_{2.5}$  when emissions are set to the “natural” value that would be observed without a scrubber, thus representing the causal effect of scrubber installation on  $\text{PM}_{2.5}$  that is “direct” in the sense that it is not attributable to changes in emissions. Formally, we defined the NDE in the multipollutant setting as  $\text{NDE} = E[Y(1; \mathbf{M}(0,0,0)) - Y(0; \mathbf{M}(0,0,0))]$ .

NIE in this context were defined to represent causal effects of scrubber installation on  $\text{PM}_{2.5}$  that are attributable only to emissions changes. In the multipollutant setting, different natural indirect effects were defined based on different multipollutant emissions. The joint natural indirect effect (JNIE) of all three mediators,  $\text{JNIE}_{123}$  (i.e., the indirect effect attributable to changes in all three emissions), was derived by subtracting the NDE from the TE,  $\text{JNIE}_{123} = \text{TE} - \text{NDE} = E[Y(1; \mathbf{M}(1,1,1)) - Y(1; \mathbf{M}(0,0,0))]$ , where TE was as defined above as the comparison between the observable potential outcomes  $Y(1; \mathbf{M}(1,1,1))$  and  $Y(0; \mathbf{M}(0,0,0))$ .

In addition to  $\text{JNIE}_{123}$ , which is of interest, we introduced a decomposition of this joint effect into the NIE attributable to changes in different combinations of the  $K = 3$  emissions. The  $\text{JNIE}_{123}$  can be decomposed into emission-specific indirect effects and the joint indirect effects of all possible pairs of emissions. See Figure 12 for a graphical representation of the various NIEs.

The mediator-specific NIE for the  $k^{\text{th}}$  emission was defined as a comparison between potential  $\text{PM}_{2.5}$  outcomes where the  $k^{\text{th}}$  emission varies between what it would be with and without a scrubber, but all other emissions are fixed to the potential value that would be observed with the scrubber. The mediator-specific NIEs for emissions of  $\text{SO}_2$ ,  $\text{NO}_x$ , and  $\text{CO}_2$  are defined as:

$$\text{NIE}_1 = E \left[ \begin{array}{l} Y(1; M_1(1), M_2(1), M_3(1)) \\ -Y(1; M_1(0), M_2(1), M_3(1)) \end{array} \right], \quad (3)$$

$$\text{NIE}_2 = E \left[ \begin{array}{l} Y(1; M_1(1), M_2(1), M_3(1)) \\ -Y(1; M_1(1), M_2(0), M_3(1)) \end{array} \right], \quad (4)$$

and

$$\text{NIE}_3 = E \left[ \begin{array}{l} Y(1; M_1(1), M_2(1), M_3(1)) \\ -Y(1; M_1(1), M_2(1), M_3(0)) \end{array} \right], \quad (5)$$

respectively.

In a similar fashion, we can define the JNIE attributable to changes in pairs of mediators  $j$  and  $k$ ,  $\text{JNIE}_{jk}$  (the second row in Figure 12). The JNIE of mediators  $j$  and  $k$  were defined as differences between the potential  $\text{PM}_{2.5}$  that would be observed with a scrubber and the analogous potential outcome but with pollutants  $j$  and  $k$  set to the values that would have been emitted without the scrubber.

Specifically, the JNIE for changes in  $\text{SO}_2$  and  $\text{NO}_x$  is defined as:

$$\text{JNIE}_{12} = E \left[ \begin{array}{l} Y(1; M_1(1), M_2(1), M_3(1)) \\ -Y(1; M_1(0), M_2(0), M_3(1)) \end{array} \right], \quad (6)$$

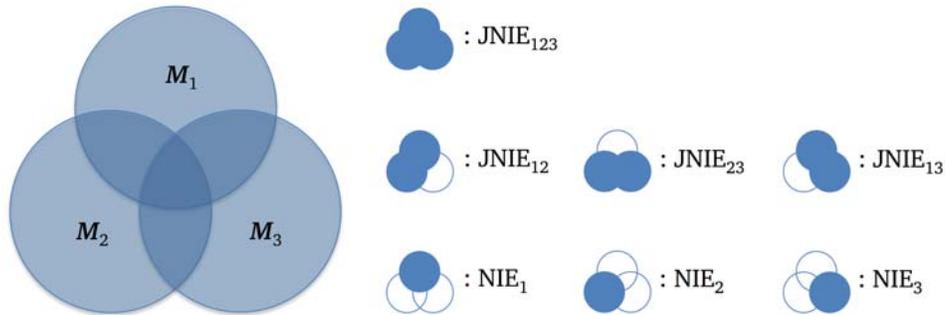


Figure 12. Graphical representation of partitioning the  $JNIE_{123}$  for three mediators. In Case Study 2, 1 =  $SO_2$ , 2 =  $NO_x$ , and 3 =  $CO_2$ .

with the analogously defined effects for ( $NO_x$  and  $CO_2$ ) and ( $SO_2$  and  $CO_2$ ) defined as

$$JNIE_{23} = E \begin{bmatrix} Y(1; M_1(1), M_2(1), M_3(1)) \\ -Y(1; M_1(1), M_2(0), M_3(0)) \end{bmatrix}, \quad (7)$$

and

$$JNIE_{13} = E \begin{bmatrix} Y(1; M_1(1), M_2(1), M_3(1)) \\ -Y(1; M_1(0), M_2(1), M_3(0)) \end{bmatrix}, \quad (8)$$

respectively.

As can be seen in Figure 12, the  $JNIE$  of each pair of mediators is not equal to the sum of corresponding mediator-specific  $NIE$ s. That is, our definitions of indirect effects did not assume additivity of effects, nor did they assume that the indirect effects were non-overlapping. This was essential in the multipollutant setting where multiple pollution emissions are measured contemporaneously and are not generally independent of each other. Thus, our methods development was in important contrast to the budding literature on mediation analysis with multiple mediators, which tends to assume non-overlapping, independent, sequential, or additive effects (Daniel et al. 2014; Imai and Yamamoto 2013; MacKinnon 2008; VanderWeele and Vansteelandt 2014).

#### Estimation: New Methods for Bayesian Nonparametric Mediation Analysis

Estimation and inference for the causal effects defined above were based on models for the joint distribution of outcomes and mediators,

$$\begin{bmatrix} Y(0; \mathbf{M}(0, 0, 0)), Y(1; \mathbf{M}(1, 1, 1)), \\ \mathbf{M}(0, 0, 0), \mathbf{M}(1, 1, 1) | \mathbf{X} \end{bmatrix}, \quad (9)$$

where  $\mathbf{X}$  denotes a vector of baseline covariates used to adjust for confounding (see Table 4). This joint distribution was not identified based on the observed data, as potential outcomes are never jointly observed in both the presence and absence of a scrubber. Thus, our estimation strategy could be characterized as consisting of three steps. First, we specified flexible Bayesian nonparametric models for the marginal distributions of observed data, which consist of values of ( $Y(0; \mathbf{M}(0,0,0)), \mathbf{M}(0,0,0)$ ) observed for power plants that did not install scrubbers, and ( $Y(1; \mathbf{M}(1,1,1)), \mathbf{M}(1,1,1)$ ) observed for power plants that did install scrubbers (Escobar and West 1995; Jara et al. 2011; Müller et al. 1996). Second, we linked all of these flexibly modeled marginal distributions into a coherent joint distribution through the use of a Gaussian copula model (Nelsen 1999). Unobserved (but observable) potential outcomes are then simulated from their posterior-predictive distributions to estimate the TE and the associative and dissociative effects. Finally, a series of assumptions, summarized in Table 5, were used to relate unobservable potential outcomes to observed relationships in the data to provide estimates of the natural direct and indirect effects. Details of the statistical models, the assumptions for identification, and the Markov chain Monte Carlo computational algorithm appear in Appendix C.

#### CASE STUDY 2: RESULTS

Table 4 indicates that the area within the 150-km radius around power plants with  $SO_2$  scrubbers installed in January 2005 had an average ambient  $PM_{2.5}$  concentration that was lower than the areas surrounding power plants without scrubbers ( $11.6 \mu\text{g}/\text{m}^3$  versus  $13.2 \mu\text{g}/\text{m}^3$ ). Plants with scrubbers emitted less  $SO_2$ , more  $NO_x$ , and more  $CO_2$  than the plants without scrubbers. However, such unadjusted comparisons were likely confounded by differences in plant characteristics, as plants with  $SO_2$  scrubbers tended to be larger (as measured by heat input), to operate

**Table 5.** Summary of Assumptions Required for Estimation of Causal Effects in Case Study 2<sup>a</sup>

Assumption	Example Interpretation	Used for	
		Principal Stratification	Mediation
Ignorability (A1)	Scrubber status is “randomized” conditional on the variables in Table 4	Yes	Yes
Gaussian Copula (A2)	The joint distribution of potential outcomes and mediators follows a Gaussian copula model	Yes	Yes
Conditional Independence I (A3)	PM <sub>2.5</sub> concentration around a plant with scrubbers is independent of what it would be without scrubbers, conditional on plant characteristics and what emissions would be with and without scrubbers	Yes	Yes
Conditional Independence II (A3★)	PM <sub>2.5</sub> concentration around a plant — under the hypothetical setting, where scrubbers are installed but emissions are fixed to what they would be without scrubbers — is independent of what it would be with scrubbers, conditional on plant characteristics of any hypothetical emissions profile (e.g., SO <sub>2</sub> with scrubbers, NO <sub>x</sub> and CO <sub>2</sub> without)	No	Yes
Distribution of Unobservable Potential Outcomes (A4)	PM <sub>2.5</sub> concentration around a plant — under the hypothetical setting, where scrubbers are installed but emissions are fixed to what they would be without scrubbers — has the same distribution of PM <sub>2.5</sub> that is observed around plants without scrubbers with that same level of emissions	No	Yes

<sup>a</sup> Details appear in Appendix C.

at a higher percent capacity (and are thus likely more efficient plants to operate), to have selective catalytic or selective noncatalytic reduction systems to control NO<sub>x</sub>, and to burn coal with a higher sulfur content.

Using the approach outlined above, having an SO<sub>2</sub> scrubber installed is estimated to cause SO<sub>2</sub> emissions to be  $-1.02 \log(\text{tons})$  lower, on average, than they would have been without the scrubber (95% posterior interval;  $-1.41, -0.68$ ). The analogous causal effects for NO<sub>x</sub> and CO<sub>2</sub> emissions were  $0.12 (-0.15, 0.37)$  and  $0.103 (-0.09, 0.29)$ , respectively, indicating that SO<sub>2</sub> scrubbers did not causally affect these emissions, on average. The TE on ambient PM<sub>2.5</sub> within a 150-km radius of a power plant was estimated to be  $-0.38 \mu\text{g}/\text{m}^3$  (95% posterior interval,  $-1.25, 0.48$ ), suggesting that having a scrubber installed causally reduced ambient PM<sub>2.5</sub> compared with what would have occurred in the absence of a scrubber, although this effect was estimated with substantial uncertainty, and was not significantly different from zero.

Principal stratification analysis can provide important evidence of the extent to which causal reductions in PM<sub>2.5</sub> are associative or dissociative with changes in emissions of SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub>. We defined average associative and dissociative effects for (1) the trio of SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub> emissions; (2) each pair of emissions, and (3) each emission individually. For the  $k^{\text{th}}$  emission, let  $\sigma_k$  denote the standard deviation of the estimated individual-level causal effect of a scrubber on the  $k^{\text{th}}$  emission, with  $\sigma_1 = 1.386$ ,  $\sigma_2 = 1.103$ , and  $\sigma_3 = 1.015$ . Let  $\sigma_{\mathcal{K}}$  denote the vector of values of  $\sigma_k$  for the pollutants in  $\mathcal{K}$ . We consider changes within  $0.5\sigma_k$  of 0 to represent little or no causal effect on emissions, and changes in excess of  $0.5\sigma_k$  to represent meaningful changes in emissions. Thus, for the EDE <sub>$\mathcal{K}$</sub> , we set  $C_{\mathcal{K}}^D = 0.5\sigma_{\mathcal{K}}$  to define the average effect of a scrubber on ambient PM<sub>2.5</sub> among power plants where there is little or no estimated causal effect on the emissions

in  $\mathcal{K}$ . For each  $\mathcal{K}$ , we defined two types of EAE:  $EAE_1$  represents the causal effect of a scrubber on  $PM_{2.5}$  among power plants where emissions were reduced by at least  $0.5\sigma_{\mathcal{K}}$  compared with what they would have been without the scrubber, and  $EAE_2$  represents the analogous effect among power plants where emissions were increased by at least  $0.5\sigma_{\mathcal{K}}$ . These correspond to  $C_{\mathcal{K}}^A = 0.5\sigma_{\mathcal{K}}$ .

For each set of emissions, Figure 13 depicts the posterior mean estimate of EDE (blue circle),  $EAE_1$  (red circle), and  $EAE_2$  (gray circle). The size of the circle is proportional to the estimated proportion of power plants that contribute to the estimate, that is, the estimated proportion of plants that have changes in emissions in accordance with the values of  $C_{\mathcal{K}}^D$  and  $C_{\mathcal{K}}^A$ . For example, the red circle in the first column of Figure 13 indicates that scrubbers are estimated to reduce  $SO_2$  emissions by at least  $C_{\mathcal{K}}^A$  for 60% of power plants, and the average causal effect on  $PM_{2.5}$  among these power plants is a reduction of  $0.60 \mu\text{g}/\text{m}^3$  compared with what would have occurred if a scrubber had not been installed. The blue circle indicates that a scrubber is estimated to have little or no causal effect on  $SO_2$  emissions among 30% of power plants, and the average causal effect of the scrubber on ambient  $PM_{2.5}$  in these plants is a reduction of  $0.25 \mu\text{g}/\text{m}^3$ . The gray circle indicates that 10% of plants are estimated to cause increased emissions of  $SO_2$  of at least  $C_{\mathcal{K}}^A$ , and the average causal effect of the scrubber on ambient  $PM_{2.5}$  around these plants is an increase of  $0.23 \mu\text{g}/\text{m}^3$ . Columns 2 and 3

of Figure 13 can be interpreted analogously, but for  $NO_x$  and  $CO_2$ , respectively. Columns 4–6 of Figure 13 depict estimates based on changes in the corresponding pairs of emissions. For example, the red circle in column 5 of Figure 13 indicates that 18% of plants were estimated to experience a reduction of  $0.51 \mu\text{g}/\text{m}^3$  in the average causal effect on  $PM_{2.5}$  among power plants where both  $SO_2$  and  $CO_2$  were reduced. Table 6 lists posterior mean and 95% posterior intervals of EDE,  $EAE_1$ , and  $EAE_2$  for all possible  $\mathcal{K}$ . Overall, we see that all estimates of EDE and  $EAE_1$  are negative, implying that scrubbers causally reduce  $PM_{2.5}$  when there is little or no change in emissions or when emissions are causally reduced. The only positive estimates in Table 6 are for  $EAE_2$  when considering increases in  $SO_2$  emissions, implying that, among plants estimated to have increased  $SO_2$  emissions, scrubbers cause increases in ambient  $PM_{2.5}$ . It should be noted that these estimates are among very few power plants, as the vast majority of plants are estimated to exhibit causal reductions in  $SO_2$ , and that these point estimates are very close to zero compared with the width of their uncertainty intervals. Also, when  $\mathcal{K}$  included  $SO_2$  emissions, estimates of  $EAE_1$  were larger in magnitude than estimates of EDE, implying that the scrubbers reduced  $PM_{2.5}$  more when there was a larger reduction in  $SO_2$  emissions. This result provides suggestive evidence of the anticipated causal pathway in which the scrubber reduced  $SO_2$  which, in turn, reduced  $PM_{2.5}$ . In contrast, when  $\mathcal{K}$  did not include  $SO_2$  emissions, estimates of EDE were similar or greater in magnitude than

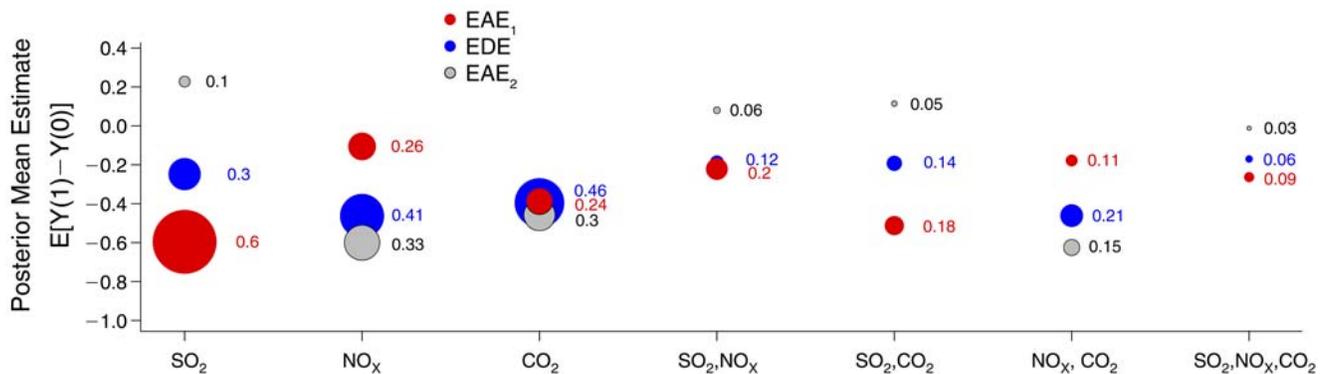


Figure 13. Posterior mean estimates of average associative (EAE) and dissociative (EDE) effects of  $SO_2$  scrubbers. Size of circle is proportional to the percent of observations estimated to belong in the corresponding strata, and the numbers are posterior mean proportions. See text for details.

**Table 6.** Posterior Mean and 95% Posterior Intervals for Expected Associative (EAE<sub>1</sub> and EAE<sub>2</sub>) and Dissociative (EDE) Effects of SO<sub>2</sub> Scrubbers

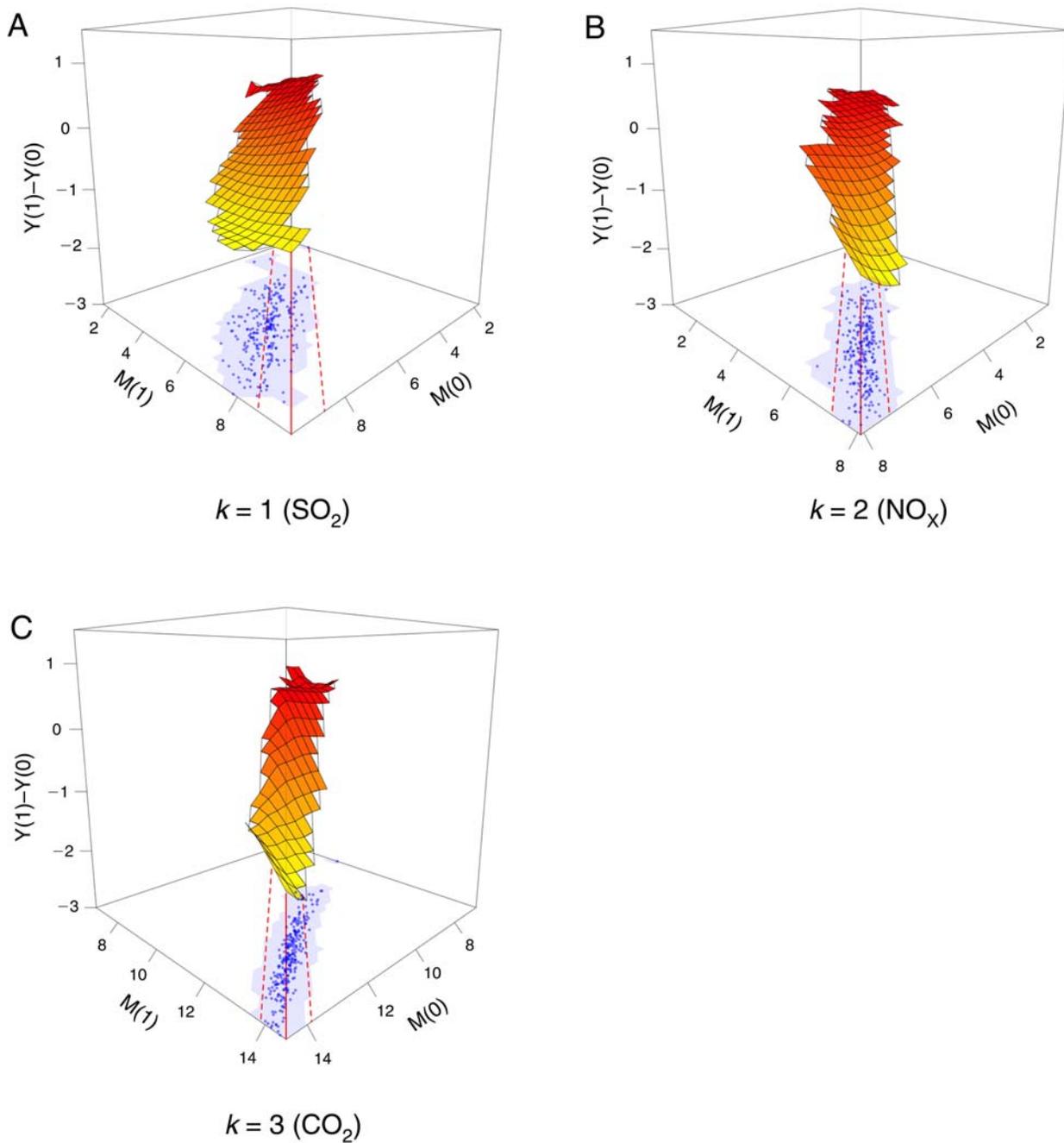
	SO <sub>2</sub> Mean (95% PI)	NO <sub>x</sub> Mean (95% PI)	CO <sub>2</sub> Mean (95% PI)	SO <sub>2</sub> & NO <sub>x</sub> Mean (95% PI)	SO <sub>2</sub> & CO <sub>2</sub> Mean (95% PI)	NO <sub>x</sub> & CO <sub>2</sub> Mean (95% PI)	SO <sub>2</sub> & NO <sub>x</sub> & CO <sub>2</sub> Mean (95% PI)
EAE <sub>1</sub>	-0.596 (-1.643 to 0.295)	-0.106 (-1.214 to 0.822)	-0.387 (-1.376 to 0.653)	-0.222 (-1.289 to 0.713)	-0.512 (-1.549 to 0.397)	-0.178 (-1.390 to 0.756)	-0.263 (-1.454 to 0.687)
EDE	-0.248 (-1.400 to 0.854)	-0.463 (-1.476 to 0.398)	-0.396 (-1.450 to 0.458)	-0.187 (-1.298 to 0.864)	-0.192 (-1.382 to 0.903)	-0.461 (-1.479 to 0.438)	-0.170 (-1.275 to 0.891)
EAE <sub>2</sub>	0.228 (-1.279 to 1.975)	-0.600 (-1.841 to 0.484)	-0.457 (-1.655 to 0.473)	0.080 (-1.448 to 1.965)	0.115 (-1.361 to 1.662)	-0.625 (-1.891 to 0.458)	-0.012 (-1.544 to 1.632)

estimates of EAE<sub>1</sub>, indicating that effects of scrubbers on PM<sub>2.5</sub> were not related to whether NO<sub>x</sub> or CO<sub>2</sub> emissions were reduced. This result suggests that NO<sub>x</sub> and CO<sub>2</sub> were not related to a causal pathway in which scrubbers reduced PM<sub>2.5</sub> through reducing these emissions. Although consistent with expectations, the above results must be interpreted in light of the wide posterior intervals, all of which included zero.

Although estimates of EDE, EAE<sub>1</sub>, and EAE<sub>2</sub> provided useful summary quantities of how the causal effect of a scrubber on PM<sub>2.5</sub> varied with the causal effect of the scrubber on emissions, we could also examine entire surfaces of how the scrubber effect on PM<sub>2.5</sub> varied with the causal effect on emissions. Figure 14 depicts three-dimensional surface plots showing how the scrubber effect on PM<sub>2.5</sub> varied across all values of  $M_k(0)$  and  $M_k(1)$  representing varying causal effects on emissions. Figure 14(A) depicts the estimated effect on PM<sub>2.5</sub> for any combination of  $(M_1(0), M_1(1))$  across the range observed in the data. Note that most of the surface lies below 0, indicating that the scrubber is estimated to reduce PM<sub>2.5</sub> for nearly all power plants. The surface is lowest in the region where  $M(1) < M(0)$ , and the steepest portion of the surface is in regions where  $M(0)$  is high; that is, the causal effect on PM<sub>2.5</sub> is estimated to be most pronounced in power plants that would have the highest emissions absent a scrubber and exhibit a causal reduction in emissions because of the scrubber. The analogous surfaces for NO<sub>x</sub> and CO<sub>2</sub> appear in Figures 14(B) and 14(C), respectively. Except for portions at only the highest values of  $M(0)$  and  $M(1)$  (i.e., the largest plants), these surfaces are much flatter in the direction of  $M_k(1) - M_k(0)$ , indicating that any causal effect of a scrubber on PM<sub>2.5</sub> varies only slightly with varying causal

effects on NO<sub>x</sub> and CO<sub>2</sub>. This is consistent with the EDE estimated to be similar (or even larger) than the estimated EAE for these emissions. Although the effect on PM<sub>2.5</sub> does not appear to vary substantially with the effects on emissions of NO<sub>x</sub> and CO<sub>2</sub>, Figure 14(B and C) depicts a downward slope in the surface in the direction of increasing  $M_k(0)$  (and  $M_k(1)$ ), indicating that the strongest scrubber effects on ambient PM<sub>2.5</sub> are evident among plants that have the highest emissions (i.e., the largest plants). This pattern is also evident from Figure 14(A). In Figure 14(A), the blue dots in the  $xy$ -plane represent the observed  $M(Z)$  and one posterior simulation of the missing  $M(1 - Z)$ . These points lie almost entirely in the region where  $M(1) < M(0)$ , reflecting that the scrubber is estimated to reduce SO<sub>2</sub> emissions for nearly all power plants. In Figure 14(B), the analogous blue dots follow more closely and symmetrically around the line  $M(1) = M(0)$ , reflecting that SO<sub>2</sub> scrubbers do not tend to affect NO<sub>x</sub> emissions. In Figure 14(C), the analogous blue dots are also tightly clustered around zero, except that the largest plants (those with high values of  $M(0)$  and  $M(1)$ ) exhibit some causal reduction in CO<sub>2</sub> emissions and the smallest plants (with low values of  $M(0)$  and  $M(1)$ ) exhibit some increase in CO<sub>2</sub> emissions.

Our overall conclusion of the principal stratification analysis was that (1) causal effects of the SO<sub>2</sub> scrubber on ambient PM<sub>2.5</sub> were most evident among plants where SO<sub>2</sub> was estimated to be causally reduced, (2) larger causal reductions in SO<sub>2</sub> were associated with larger causal reductions in PM<sub>2.5</sub> regardless of changes in other emissions, and (3) causal reductions in NO<sub>x</sub> and CO<sub>2</sub> were not associated with causal reductions in PM<sub>2.5</sub>. Although not conclusive about the mediated indirect effects whereby the scrubber reduced PM<sub>2.5</sub> through reducing SO<sub>2</sub> emissions,



**Figure 14.** Average surface plots of the causal effect on  $\text{PM}_{2.5}$  for different values of  $(M_k(0), M_k(1))$ . Values of  $(M_k(0), M_k(1))$  are plotted on the  $x$  and  $y$  axes, and determine the causal effect of a scrubber on emission  $k$ . The corresponding value of the causal effect of a scrubber on  $\text{PM}_{2.5}$  ( $Y(1) - Y(0)$ ) is plotted on the  $z$  axis. The blue cloud of points are simulations of  $(M_k(0), M_k(1))$  for one Markov chain Monte Carlo iteration to represent the range of values of  $(M_k(0), M_k(1))$  consistent with the observed data. Red lines are at  $M_k(0) = M_k(1)$  (solid line) and  $\pm\sigma_k$  (dashed lines). (A)  $k = 1$  ( $\text{SO}_2$ ). (B)  $k = 2$  ( $\text{NO}_x$ ). (C)  $k = 3$  ( $\text{CO}_2$ ).

this analysis was consistent with the presence of such a causal pathway, although estimates had large uncertainty.

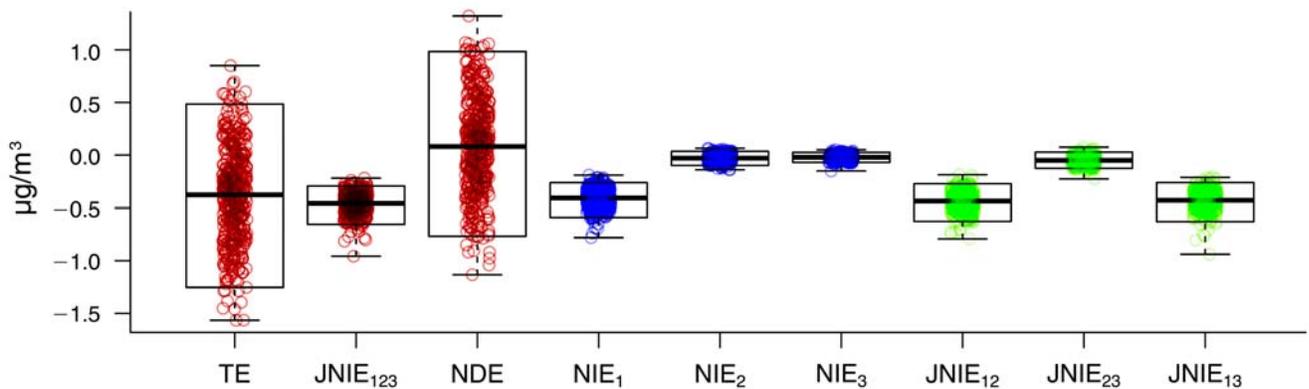
Augmenting the principal stratification analysis with assumptions about unobservable potential outcomes that conceive of independent manipulations of scrubbers and each emission individually made estimation of natural direct and indirect effects possible. These effects speak more directly to the extent to which the effect of an SO<sub>2</sub> scrubber on ambient PM<sub>2.5</sub> within 150 km of a power plant is mediated through various emissions. Again, the mediator-specific NIE of the *k*<sup>th</sup> emission and the JNIE of the *j*<sup>th</sup> and *k*<sup>th</sup> emissions are denoted as NIE<sub>*k*</sub> and JNIE<sub>*jk*</sub>, respectively, where *k* = 1 indicates SO<sub>2</sub>, *k* = 2 indicates NO<sub>x</sub>, and *k* = 3 indicates CO<sub>2</sub>.

Table 7 summarizes point estimates and 95% posterior intervals of the TE, NDE, JNIE<sub>123</sub>, JNIE<sub>12</sub>, JNIE<sub>23</sub>, JNIE<sub>13</sub>, and the individual NIEs. Figure 15 depicts boxplots of the entire posterior distributions of these quantities. The NDE posterior mean estimate was a reduction in ambient PM<sub>2.5</sub> concentration of 0.080 µg/m<sup>3</sup> (95% posterior interval; -0.77, 0.98) and represented the direct effect of a scrubber on ambient PM<sub>2.5</sub> that is not mediated through any emissions changes. In other words, the NDE represents the amount that PM<sub>2.5</sub> would decrease if a scrubber were installed but emissions of SO<sub>2</sub>, NO<sub>x</sub>, and CO<sub>2</sub> were somehow fixed to remain constant at what they would have been without the scrubber, and was estimated to be very close to zero. The indirect effect via all three

**Table 7.** Posterior Mean and 95% Posterior Intervals for Total, Direct, and Indirect Effects of SO<sub>2</sub> Scrubbers<sup>a</sup>

	TE	JNIE <sub>123</sub>	NDE
Mean	-0.379	-0.459	0.080
95% P.I.	(-1.253 to 0.484)	(-0.654 to -0.295)	(-0.767 to 0.984)
	NIE <sub>1</sub>	NIE <sub>2</sub>	NIE <sub>3</sub>
Mean	-0.408	-0.030	-0.021
95% P.I.	(-0.589 to -0.262)	(-0.099 to 0.035)	(-0.070 to 0.025)
	JNIE <sub>12</sub>	JNIE <sub>23</sub>	JNIE <sub>13</sub>
Mean	-0.437	-0.051	-0.429
95% P.I.	(-0.626 to -0.272)	(-0.127 to 0.028)	(-0.629 to -0.261)

<sup>a</sup> TE = total effect; NDE = natural direct effect; NIE = natural indirect effect; and JNIE = joint natural indirect effect. Subscripts represent *k* emissions, where 1 = SO<sub>2</sub>; 2 = NO<sub>x</sub>; and 3 = CO<sub>2</sub>.



**Figure 15.** Posterior distributions of direct and indirect effects in the analysis of SO<sub>2</sub> scrubbers. TE = total effects, NDE = natural direct effect, NIE = natural indirect effect, and JNIE = joint natural indirect effect. Subscripts represent *k* emissions, where 1 = SO<sub>2</sub>, 2 = NO<sub>x</sub>, and 3 = CO<sub>2</sub>.

emissions ( $JNIE_{123}$ ) was estimated with posterior mean  $-0.46$  (95% posterior interval;  $-0.65, -0.30$ ), which represents the reduction in  $PM_{2.5}$  that would occur around a plant with a scrubber compared with what would happen if emissions of  $SO_2$ ,  $NO_x$ , and  $CO_2$  were somehow changed to what they would have been absent the scrubber. The relative magnitudes of the TE and  $JNIE_{123}$  indicated that virtually all of the TE of a scrubber on ambient  $PM_{2.5}$  is jointly mediated through changes in  $SO_2$ ,  $NO_x$ , and  $CO_2$ . The NIE for  $SO_2$  ( $NIE_1$ ) was estimated with posterior mean of  $-0.41$  (95% posterior interval;  $-0.59, -0.26$ ), indicating that most of the joint indirect effect could be attributed to reductions in  $SO_2$ . Posterior mean estimates of NIE for  $NO_x$  ( $NIE_2$ ) and the NIE for  $CO_2$  ( $NIE_3$ ) were both very close to zero —  $-0.030$  (95% posterior interval;  $-0.10, 0.035$ ) for  $NIE_2$  and  $0.02$  (95% posterior interval;  $-0.07, 0.03$ ) for  $NIE_3$ . Estimates of the joint indirect effects that involved  $SO_2$  ( $JNIE_{12}$  and  $JNIE_{13}$ ) were close in magnitude to that of  $JNIE_{123}$  and  $NIE_1$ , indicating that combining causal reductions of  $SO_2$  with causal reductions of either  $NO_x$  or  $CO_2$  did not substantially change the mediated effect more than did the reductions in  $SO_2$  alone.

The overall conclusion of the causal mediation analysis was that (1)  $SO_2$  scrubbers appeared to causally reduce  $PM_{2.5}$  on average, although this effect was estimated with large uncertainty and not significantly different from zero; (2) any effect of scrubbers on  $PM_{2.5}$  was almost entirely mediated through causal reductions in  $SO_2$ ; (3)  $SO_2$  scrubbers did not causally affect  $NO_x$  or  $CO_2$  emissions, on average; and (4) there was no evidence that effects of scrubbers on  $PM_{2.5}$  were mediated through changes in  $NO_x$  or  $CO_2$ . Appendix C (section C.5) provides an examination of the extent to which the indirect effects of the three emissions overlap one another. Appendix D, available on the Web, presents a sensitivity analysis analogous to the analysis presented here, but for a data set that links power plants to all monitors within a radius of 75 km instead of 150 km. The results of the analysis with 75-km data linkage were very similar to those presented here.

## CONCLUSION AND DISCUSSION OF CASE STUDY 2

In this case study, we evaluated the effectiveness of a specific regulatory intervention — the presence of an  $SO_2$  scrubber on a coal-fired power plant in 2005 — in terms of (1) the intervention's causal effect on annual emission of  $SO_2$ ,  $NO_x$ , and  $CO_2$  in 2005, (2) the intervention's causal effect on average annual ambient  $PM_{2.5}$  in 2005 among monitors within 150 km of a power plant, and (3) the extent to which the intervention's causal effect on ambient  $PM_{2.5}$  is mediated through causal reductions in multiple emissions. We focused on the method of causal mediation

analysis to provide the first empirical evidence of the presumed causal relationships that motivate emissions control interventions, which continue to be important strategies for improving ambient air quality and, ultimately, human health. Given that our questions of interest — and indeed many accountability questions — pertain to mediated effects of multiple pollutants, which are measured contemporaneously and possibly interact with one another, we developed new methods for principal stratification and causal mediation analysis for multiple, contemporaneous, and non-independent mediators. We introduced Bayesian nonparametric modeling and estimation techniques to provide flexible models for studying the observed data and linked observed data distributions into joint distributions of potential outcomes using explicit and transparent assumptions (presented in detail in section C.3 of Appendix C). The results of our analysis were largely consistent with expectations:  $SO_2$  scrubbers appeared to causally reduce  $SO_2$  emissions and ambient  $PM_{2.5}$  (within 150 km), and the causal effect on  $PM_{2.5}$  was mediated almost entirely by causal reductions in  $SO_2$  emissions and not through alterations in  $NO_x$  or  $CO_2$  emissions.

The results of the principal stratification and causal mediation analyses should be interpreted jointly in support of our conclusions. For the principal stratification analysis, the difference between associative and dissociative effects was most pronounced when considering emissions of  $SO_2$ , either individually or in combination with other emissions. This pronounced difference indicated that power plants that exhibited large causal effects on  $SO_2$  emissions also exhibited large effects on ambient  $PM_{2.5}$ . For all other emissions, estimates of EDE were similar to or larger than estimates of EAE, suggesting that scrubbers affected  $PM_{2.5}$  through pathways that did not involve these emissions (although none of these estimates differed significantly from zero). From the mediation analysis, estimates of the NIE of  $SO_2$  and the JNIEs involving  $SO_2$  ( $JNIE_{12}$ ,  $JNIE_{13}$ , and  $JNIE_{123}$ ) were all significantly less than zero and similar in magnitude to one another, highlighting mediation of the effect of scrubbers on  $PM_{2.5}$  attributable to reductions in  $SO_2$  emissions. All other NIE estimates not involving  $SO_2$ , as well as estimates of the NDE, were close to zero, indicating that scrubber-induced reductions in  $PM_{2.5}$  can be attributed almost exclusively to causal reductions in  $SO_2$  emissions.

A key feature of Case Study 2 is the integration of a principal stratification analysis and a causal mediation analysis — both of which rely on the same modeling assumptions for the observed data. A key difference between these two analyses pertains to the presence or absence of assumptions about potential outcomes that are unobservable for every

observation in the study sample. Thus, we begin with a principal stratification analysis relying on assumptions for the observable outcomes  $Y(z, \mathbf{M}(z, z, z)), \mathbf{M}(z, z, z)$  to identify the principal causal effects, and then augment these assumptions with assumptions about unobservable potential outcomes [e.g.,  $Y(1, \mathbf{M}(0, 0, 0))$ ] to estimate mediation effects. The explicit connection between principal stratification and causal mediation analyses explored here represents, to our knowledge, the most comprehensive consideration of these two approaches and the implications of their results in the context of a single analysis.

Interpretation of the results of this case study should be viewed in light of several important limitations related to key analytic choices and assumptions. First, is the relative simplicity with which we linked power plants to monitors. Specifically, our strategy simply linked power plants to all of the ambient monitors within 150 km, resulting in an analysis that likely did not reflect the full effect of emissions changes from power plants on ambient air quality, which are expected to have implications at distances much greater than 150 km. Furthermore, monitors located within 150 km of more than one power plant were linked only to the closest power plant, which resulted in many power plants not being assigned to any monitor and thus being excluded from our analysis. As a result, our analysis assumed that impacts on ambient  $\text{PM}_{2.5}$  attributable to emissions changes in these excluded power plants were distributed evenly across areas surrounding the scrubber and no scrubber power plants included in the analysis. More sophisticated strategies to link ambient monitors to power plants, based on features such as atmospheric conditions and weather patterns, are warranted, but analysis of the data constructed here represented an important approximation that still yielded valuable inferences, especially with respect to quantifying causal pathways.

A second limitation of this analysis was that it assumed that the factors listed in Table 4 were sufficient to control for confounding, which in this case would consist of differences between power plants or other features related to ambient  $\text{PM}_{2.5}$  that are also associated with whether a power plant had a scrubber installed in 2005. Importantly, we controlled for factors such as the size of the power plant (as measured by heat input), the sulfur content of the coal used, concurrent  $\text{NO}_x$  controls installed (including whether the plant had a selective catalytic reduction or selective non-catalytic reduction system), plant operating time, the phase of participation in the ARP, and the percentage of full capacity at which the plant operated (as a proxy of plant operating efficiency). Nonetheless, the prospect of confounding attributed to other factors remains, and more work is warranted to incorporate further information

on plant characteristics, information about other emissions (e.g., primary particles), or features of secondary formation of  $\text{PM}_{2.5}$  not directly captured by direct emissions.

A third limitation was that we considered analysis of a single year and regarded a power plant as “treated” if it had a scrubber installed in January 2005, without regard to how long the scrubber had been installed or to changes in emissions and ambient  $\text{PM}_{2.5}$  over time. Future work will develop a framework to accommodate longitudinal analysis by using Bayesian dynamic models, which could update information from the past and smooth the effects over the course of a several years (Kim et al. 2015). A related limitation of our cross-sectional design is that it compresses significant seasonal variation in secondary formation of  $\text{PM}_{2.5}$  into a single annual measure. Separate analyses for different seasons would help shed light on whether the impact of scrubbers on  $\text{PM}_{2.5}$  varied with different types of secondary formation.

Despite the limitations of this second case study, we have conducted the first empirical investigation of the presumed causal pathways that have prompted a variety of air quality control strategies aimed at reducing harmful emissions from power plants on a national scale. Using a principled causal inference framework and rigorous analysis to quantify causal pathways, we evaluated the effectiveness of scrubber installation for reducing emissions and ambient  $\text{PM}_{2.5}$ , representing an analysis of two important links in the chain of accountability amid the realities of actual regulatory implementation. The health implications of our analysis rely on the presumed link between ambient  $\text{PM}_{2.5}$  and health outcomes, but the methods presented here can be applied in other multipollutant accountability settings, including extensions of the current analysis to investigate, for example, the extent to which reductions in multipollutant emissions mediate causal health effects or the extent to which scrubber-induced changes in ambient  $\text{PM}_{2.5}$  (or other pollutants) mediate causal effects on health outcomes.

### Extension of Case Study 2 to Health-Outcomes Analyses

The impact of emissions from any given source on ambient air quality involves complicated wind patterns, geography, topography, atmospheric conditions, and myriad other factors that determine pollution transport. A key limitation of our analysis was that we did not specifically consider pollution transport beyond the proximity-based strategy of linking each power plant to all monitors within a specified radius (and excluding many power plants that were not linked in this process). As a consequence, our analysis, which was limited to the effects of scrubbers on ambient  $\text{PM}_{2.5}$  within 150 km of a monitor,

may have missed many important consequences of controlling emissions at power plants. The potential impact of failure to account for these consequences was evident in the wide posterior intervals for the estimated total effect of scrubbers on  $PM_{2.5}$ .

In addition to being an important limitation of the analysis of Case Study 2, we have found that this failure to account for pollution transport practically prohibited extension of our multipollutant analysis to the study of scrubber installation on health outcomes. One factor is that the necessary extra step of linking emissions and ambient air quality data to health outcomes data entails its own limitations. A more important factor is the scientific reality that any possible impact of scrubber installation on health outcomes is expected to be small in comparison with the effects on emissions and ambient  $PM_{2.5}$ , which magnifies the limitation of failing to account for pollution transport. In short, the modulation of signal as the analysis moves down the chain of accountability amplifies the limitations of the present analysis, which was restricted to emissions and ambient air quality.

Increasing the sophistication of the use of important atmospheric and geographic data by means of our newly developed statistical methods is an important avenue for future research that we continue to explore. We have conducted preliminary analyses of health outcomes using the methodology presented in this report, but scaled back to a single pollutant. We conducted these preliminary investigations with two different data-linkage processes, which are described in detail in Appendix E. The first process is an extension of the data linkage described in the analysis of Case Study 2, whereby power plants were linked to monitors located within 150 km and the Medicare health outcomes were linked from all zip codes within 6 miles of a linked monitoring station. The process resulted in a data set for which the power plant is the unit of analysis,  $PM_{2.5}$  is the average of monitored values from stations within a 150-km radius, and Medicare health outcomes are aggregated among all Medicare beneficiaries residing within 6 miles of a monitoring station that, in turn, is at most 150 km from the power plant.

As an alternative process of linking data, we explored a monitor-centered linkage, in which the ambient monitoring location is the unit of analysis. For this linkage, Medicare health outcomes at all zip codes located within 6 miles of a monitor were aggregated to the monitor. Aggregation of power-plant data was achieved by calculating, for each monitoring location, a weighted average of measures from all surrounding power plants, with weights determined by the distance between the monitor and each power plant. This allowed every power plant to contribute

information to every monitor, with the closest power plants contributing the most and very distant power plants contributing virtually no information.

For both of these data-linkage processes, we conducted preliminary mediation analyses to investigate the effect of  $SO_2$  scrubbers on Medicare health outcomes and the extent of mediation through causal effects on ambient  $PM_{2.5}$ . We found the health-outcomes analyses to exhibit wide uncertainty intervals and high sensitivity to the way in which the data on ambient monitoring data, power plants, health outcomes, and confounders were linked. Thus, with the methods we have explored to date, we judged the preliminary analyses of health outcomes to lack sufficient rigor to include in this report and have only considered them to explore the existing limitations and ways to refine the analysis to one that could be used for accountability.

### Additional Work in Progress

In addition to the analysis of Case Study 2, we are conducting a variety of ongoing analyses of causal impacts of power-plant-emissions controls. We have rigorously evaluated the causal impact of  $SO_2$  and  $NO_x$  control strategies on  $SO_2$  and  $NO_x$  emissions among 995 coal-burning EGUs during the years 1997–2012. We are also continuing to extend the analysis of Case Study 2 to investigate the extent to which the causal effect of  $SO_2$  scrubber installation on Medicare health outcomes is mediated through reductions in ambient  $PM_{2.5}$ , as described in the previous section. In particular, we are developing more sophisticated procedures for linking data between power plants, ambient monitors, and residential zip codes of Medicare beneficiaries. We are also exploring the development of new statistical methods for causal inference with interference to incorporate the nature of pollution transport into causal estimates.

---

## DISCUSSION AND CONCLUSIONS

---

Over the past ten years, important progress in accountability assessment has brought a new dimension to the scientific evidence available for informing policy decisions. Although important challenges remain, the perspectives and methods in this report represent progress towards rigorous evaluation of large-scale regulatory policies. Sharpening the distinction between analytic perspectives for exposure–response estimation and for estimating causal effects of well-defined actions is necessary in order to advance accountability assessment beyond evaluation of localized, abrupt actions and towards informing policy debates with evidence of the effects of broad and complex

regulations involving action over long time frames. We have outlined the particular relevance of potential-outcomes methods for causal inference for advancing the goals of accountability assessment in order to focus on the direct evaluation of the effectiveness of specific policies or actions. Our analysis in Case Study 1 illustrated how potential-outcomes reasoning can be deployed towards the goals of long-term direct-accountability assessment. Our analysis in Case Study 2 outlined the development of new statistical methods for multipollutant accountability assessment and illustrated how potential-outcomes perspectives can be useful for quantifying various causal pathways through which an air quality intervention impacts outcomes.

The deployment of potential-outcomes methods for direct-accountability assessment represents an important new direction for accountability research and, more broadly, for air pollution epidemiology. Defining causal consequences of well-defined actions — which we refer to as direct accountability — stands in marked contrast to studying the associations between exposures and the onset of clinical disease. The analytic perspectives and associated statistical methods presented here are consistent with a recent emphasis on consequentialist epidemiology, which shifts focus away from identifying underlying causes of disease and towards development of consequential interventions (Galea 2013). This change in focus is not intended to diminish the importance of the vast array of air pollution epidemiological evidence that motivated the need to intervene in order to control population exposures, but rather to emphasize the need to provide equally strong evidence of the consequences of specific intervention strategies that aim to protect public health and the environment. Although no single analytic strategy can overcome all the challenges inherent in accountability assessment, the best science should be generated from a variety of available approaches. We argue that rigorous efforts to evaluate directly causal effects of well-defined regulatory interventions constitute one such approach that, while distinct from traditional epidemiological tools, is essential in the current regulatory climate.

---

### ACKNOWLEDGMENTS

We wish to acknowledge in particular the contributions to this report made by John Bachmann, whose wealth of air quality regulatory knowledge improved our analysis and understanding of both case studies. He helped us to ensure that the case studies used here were consistent with the realities of air quality management in the United States over the past several decades. Interpreting preliminary

results of the case studies with his aid led to significant improvements in subsequent analyses. We also wish to acknowledge George Allen and Paul Miller for helpful insights pertaining to the analysis of Case Study 1, and Petros Koutrakis and Joel Schwartz for helpful discussions about the analysis of Case Study 2.

---

### REFERENCES

- Auffhammer M, Bento AM, Lowe SE. 2009. Measuring the effects of the Clean Air Act Amendments on ambient PM<sub>10</sub> concentrations: the critical importance of a spatially disaggregated analysis. *J Environ Econ Manage* 58:15–26.
- Banzhaf HS, Burtraw D, Evans D, Krupnick A. 2006. Valuation of natural resource improvements in the Adirondacks. *Land Economics* 82:445–464.
- Buonocore JJ, Dong X, Spengler JD, Fu JS, Levy JI. 2014. Using the Community Multiscale Air Quality (CMAQ) model to estimate public health impacts of PM<sub>2.5</sub> from individual power plants. *Environ Int* 68:200–208.
- Burtraw D. 1999. Cost savings, market performance and economic benefits of the US Acid Rain Program. In: *Pollution for Sale: Emissions Trading and Joint Implementation* (Sorrell S, Skea J, eds). Cheltenham, UK:Edward Elgar, 43–62.
- Burtraw D, Krupnick A, Mansur E, Austin D, Farrell D. 1998. Costs and benefits of reducing air pollutants related to acid rain. *Contemp Econ Policy* 16:379–400.
- Cefalu M, Zigler CM. 2015. Posterior-predictive treatment assignments and the estimation of causal effects. Technical Report.
- Chay K, Dobkin C, Greenstone M. 2003. The Clean Air Act of 1970 and adult mortality. *J Risk Uncertain* 27:279–300.
- Chen Y, Ebenstein A, Greenstone M, Li H. 2013. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proc Natl Acad Sci USA* 110:12936–12941.
- Chestnut LG, Mills DM. 2005. A fresh look at the benefits and costs of the US acid rain program. *J Environ Manage* 77:252–266.
- 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.
- Correia AW, Pope CA III, Dockery DW, Wang Y, Ezzati M, Dominici F. 2013. Effect of air pollution control on life expectancy in the United States: an analysis of 545 U.S.

- counties for the period from 2000 to 2007. *Epidemiology* 24:23–31.
- Crosas M. 2011. The Dataverse Network: an open-source application for sharing, discovering and preserving data. *D-Lib Magazine* 17.
- Crump RK, Hotz VJ, Imbens GW, Mitnik OA. 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96:187–199.
- Currie J, Walker R. 2011. Traffic congestion and infant health: evidence from E-ZPass. *Am Econ J Appl Econ* 3:65–90.
- Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. 2014. Causal mediation analysis with multiple mediators. *Biometrics*; doi: 10.1111/biom.12248 [Online 28 Oct 2014].
- Deschenes O, Greenstone M, Shapiro JS. 2012. Defensive Investments and the Demand for Air Quality: Evidence from the NO<sub>x</sub> Budget Program and Ozone Reductions. Working Paper 18267. Cambridge, MA:National Bureau of Economic Research.
- Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, et al. 1993. An association between air pollution and mortality in six US cities. *N Engl J Med* 329:1753–1759.
- Dominici F, Greenstone M, Sunstein CR. 2014. Particulate matter matters. *Science* 344:257–259.
- Dominici F, Peng RD, Bell ML, Pham L, McDermott A, Zeger SL, et al. 2006. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA* 295:1127–1134.
- Escobar MD, West M. 1995. Bayesian density estimation and inference using mixtures. *J Am Stat Assoc* 90:577–588.
- Frangakis CE, Rubin DB. 2002. Principal stratification in causal inference. *Biometrics* 58:21–29.
- 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. *JAMA* 285:897–905.
- Galea S. 2013. An argument for a consequentialist epidemiology. *Am J Epidemiol*; doi: 10.1093/aje/kwt172 [Online 10 September 2013].
- Glass TA, Goodman SN, Hernn MA, Samet JM. 2013. Causal inference in public health. *Annu Rev Public Health* 34:61–75.
- Greenstone M. 2004. Did the Clean Air Act cause the remarkable decline in sulfur dioxide concentrations? *J Environ Econ Manage* 47:585–611.
- Greenstone M, Gayer T. 2009. Quasi-experimental and experimental approaches to environmental economics. *J Environ Econ Manage* 57:21–44.
- Hedley AJ, Wong C-M, Thach TQ, Ma S, Lam T-H, 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.
- HEI Accountability Working Group. 2003. Assessing the Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research. Communication 11. Boston, MA:Health Effects Institute.
- Health Effects Institute. 2010. Proceedings of an HEI workshop on further research to assess the health impacts of actions taken to improve air quality. Communication 15. Boston, MA:Health Effects Institute.
- Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. 2008. Observational studies analyzed like randomized experiments. *Epidemiology* 19:766–779.
- Hidy G, Brook J, Demerjian K, Pennell W, Scheffe R. 2011. Technical Challenges of Multipollutant Air Quality Management. New York:Springer.
- Ho DE, Imai K, King G, Stuart EA. 2007. Matching as non-parametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15:199–236.
- Hubbell B. 2012. Assessing the results of air quality management programs. *EM Magazine*, October 2015: 8–15.
- Imai K, Yamamoto T. 2013. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. *Political Analysis* 21:141–171.
- Jara A, Hanson TE, Quintana FA, Müller P, Rosner GL. 2011. DPpackage: Bayesian semi- and nonparametric modeling in R. *J Stat Softw* 40(5):1–30.
- Joffe MM, Greene T. 2009. Related causal frameworks for surrogate outcomes. *Biometrics* 65:530–538.
- Kim C, Daniels MJ, Roy JA, Marcus BH. 2015. Longitudinal Causal Mediation Analysis for Behavioral Trials. Technical Report.

- King G. 2007. An introduction to the DataVerse Network as an infrastructure for data sharing. *Sociol Methods Res* 36:173–199.
- King G, Zeng L. 2006. The dangers of extreme counterfactuals. *Political Analysis* 14:131–159.
- Krewski D, Burnett R, Goldberg M, Hoover BK, Siemiatycki J, Jerrett M, et al. 2003. Overview of the reanalysis of the Harvard Six Cities Study and American Cancer Society study of particulate air pollution and mortality. *J Toxicol Environ Health Part A* 66:1507–1552.
- Laden F, Schwartz J, Speizer FE, Dockery DW. 2006. Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities Study. *Am J Respir Crit Care Med* 173(6):667–672
- Levy JI, Baxter LK, Schwartz J. 2009. Uncertainty and variability in health-related damages from coal-fired power plants in the United States. *Risk Analysis* 29:1000–1014.
- Levy JI, Wilson AM, Zwack LM. 2007. Quantifying the efficiency and equity implications of power plant air pollution control strategies in the United States. *Environ Health Perspect* 115(5):743–750.
- MacKinnon DP. 2008. *Introduction to Statistical Mediation Analysis*. New York:Routledge.
- Moore K, Neugebauer R, Lurmann F, Hall J, Brajer V, Alcorn S, Tager I. 2010. Ambient ozone concentrations and cardiac mortality in Southern California 1983–2000: application of a new marginal structural model approach. *Am J Epidemiol* 171:1233–1243.
- Müller P, Erkanli A, West M. 1996. Bayesian curve fitting using multivariate normal mixtures. *Biometrika* 83:67–79.
- National Research Council. 1992. *Rethinking the Ozone Problem in Urban and Regional Air Pollution*. Photocopy edition. Washington, DC:National Academies Press.
- National Research Council. 2004. *Air Quality Management in the United States*. Washington, DC: National Academies Press.
- Nelsen RB. 1999. *An Introduction to Copulas*. Lecture Notes in Statistics, vol. 139. New York:Springer.
- Pearl J. 2001. Direct and indirect effects. In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 411–420.
- Pearl J. 2011. Principal stratification — a goal or a tool? *Int J Biostat* 7:20.
- Peng RD, Chang HH, Bell ML, McDermott A, Zeger SL, Samet JM, et al. 2008. Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among Medicare patients. *JAMA* 299:2172–2179.
- Pope CA III. 1996. Particulate pollution and health: a review of the Utah valley experience. *J Expo Anal Environ Epidemiol* 6:23–34.
- Pope CA III, Ezzati M, Dockery DW. 2009. Fine-particulate air pollution and life expectancy in the United States. *New Engl J Med* 360:376–386.
- Pope CA III, Rodermund DL, Gee MM. 2007. Mortality effects of a copper smelter strike and reduced ambient sulfate particulate matter air pollution. *Environ Health Perspect* 115:679–683.
- Rich DQ, Kipen HM, Huang W, Wang G, Wang Y, Zhu P, et al. 2012. Association between changes in air pollution levels during the Beijing Olympics and biomarkers of inflammation and thrombosis in healthy young adults. *JAMA* 307:2068–2078.
- Robins JM, Greenland S. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3:143–155.
- Robins JM, Hernan M, Brumback B. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550.
- Rosenbaum PR. 1984. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *J R Stat Soc Ser A (General)* 147:656–666.
- Rosenbaum PR, Rubin DB 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55.
- Rubin DB. 1978. Bayesian inference for causal effects: the role of randomization. *Ann Stat* 6:34–58.
- Rubin DB. 2004. Direct and indirect causal effects via potential outcomes. *Scand Stat Theory Appl* 31:161–170.
- Rubin DB. 2008. For objective causal inference, design trumps analysis. *Ann Appl Stat* 2:808–840.
- Samet JM. 2011. The Clean Air Act and health: a clearer view from 2011. *New Engl J Med* 365:198–201.
- Schmalensee R, Stavins R. 2012. *The SO<sub>2</sub> Allowance Trading System: the ironic history of a grand policy experiment*. Technical report. Cambridge, MA:National Bureau of Economic Research.

- Stuart EA. 2010. Matching methods for causal inference: a review and a look forward. *Stat Sci* 25:1–21.
- Tonne C, Beevers S, Armstrong B, Kelly F, Wilkinson P. 2008. Air pollution and mortality benefits of the London Congestion Charge: spatial and socioeconomic inequalities. *Occup Environ Med* 65:620–627.
- U.S. Environmental Protection Agency. 2009. Integrated Science Assessment for Particulate Matter. Research Triangle Park, NC:U.S. Environmental Protection Agency Office of Research and Development.
- U.S. Environmental Protection Agency. 2010. The Benefits and Costs of the Clean Air Act: 1990 to 2020. Revised Draft Report. Research Triangle Park, NC:U.S. Environmental Protection Agency Office of Air and Radiation.
- U.S. Environmental Protection Agency. 2011. National Emissions Inventory (NEI) Air Pollutant Emissions Trends Data: 1970–2011. Available: [www.epa.gov/ttn/chieftrends/index.html](http://www.epa.gov/ttn/chieftrends/index.html).
- U.S. Environmental Protection Agency. 2012. Environmental Benefits Mapping and Analysis Program (BenMAP) User's Manual. Technical report. Research Triangle Park, NC:U.S. Environmental Protection Agency.
- U.S. Environmental Protection Agency. 2013. Workshop on designing research to assess air quality and health outcomes from air pollution regulations. In: *Designing Research to Assess Air Quality and Health Outcomes from Air Pollution Regulations*. Research Triangle Park, NC:U.S. Environmental Protection Agency.
- 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 Magazine* October 2012:22–28.
- VanderWeele TJ. 2009. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* 20:18–26.
- VanderWeele TJ. 2008. Simple relations between principal stratification and direct and indirect effects. *Stat Probabil Letters* 78:2957–2962.
- VanderWeele T, Vansteelandt S. 2009. Conceptual issues concerning mediation, interventions and composition. *Stat Interface* 2:457–468.
- VanderWeele T, Vansteelandt S. 2014. Mediation analysis with multiple mediators. *Epidemiol Method* 2:95–115.
- Zanobetti A, Dominici F, Wang Y, Schwartz JD. 2014. A national case-crossover analysis of the short-term effect of PM<sub>2.5</sub> on hospitalizations and mortality in subjects with diabetes and neurological disorders. *Environ Health* 13:38.
- Zeger SL, Dominici F, McDermott A, Samet JM. 2008. Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000–2005). *Environ Health Perspect* 116:1614–1619.
- Zigler CM, Dominici F. 2014. Point: clarifying policy evidence with potential-outcomes thinking beyond exposure–response estimation in air pollution epidemiology. *Am J Epidemiol* 180:1133–1140.
- Zigler CM, Dominici F, Wang Y. 2012. Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics (Oxford, England)* 13:289–302.

---

#### HEI QUALITY ASSURANCE STATEMENT

---

The conduct of this study was subjected to independent audits by Mr. David Bush of T&B Systems, Inc. Mr. Bush is an expert in quality assurance for air quality monitoring studies and data management. The audits included an on-site review of study activities for conformance to the study protocol and operating procedures. The dates of the audits are listed below with the phase of the study examined.

#### January 8–9, 2014

An on-site audit was conducted at the Harvard School of Public Health, Boston, MA. Mr. Charles Blanchard assisted with the audit. The audit concentrated on the study's analytical and data management activities, and included an audit of the study's database. Several data points were traced through the entire data processing sequence to verify the integrity of the database. Recommendations resulting from the audit primarily concentrated on correcting minor issues associated with the calculations of some of the air quality metrics.

#### February 2016

The final report was reviewed to verify that issues identified during the January 2014 audit had been address. All issues had been addressed by the authors. The finalized data sets were reviewed and compared against source data, with no significant issues noted.

Written reports of each inspection were provided to the HEI project manager, who transmitted the findings to the Principal Investigator. These quality assurance audits demonstrated that the study was conducted by an experienced team with a high concern for data quality. Study per-

sonnel were very responsive to audit recommendations, providing formal responses that adequately addressed all issues. The report appears to be an accurate representation of the study.



David H. Bush, Quality Assurance Officer

---

### MATERIALS AVAILABLE ON THE WEB

---

Appendices A through E contain supplemental material not included in the printed report. They are available on the HEI Web site <http://pubs.healtheffects.org>.

Appendix A. Technical Details for Case Study 1: PM<sub>10</sub> Nonattainment Designations

Appendix B. Sensitivity Analysis to the Pruning of Observations in Case Study 1

Appendix C. Technical Details for Case Study 2: Power-Plant Emissions Controls

Appendix D. Results from the Power-Plant Case Study with 75-km Data Linkage

Appendix E. Preliminary Extension of the Power-Plant Case Study to Health Outcomes

---

### ABOUT THE AUTHORS

---

**Corwin Matthew Zigler**, Ph.D., is an assistant professor of biostatistics in the Department of Biostatistics at the Harvard T.H. Chan School of Public Health. His research focus is the development of statistical methods for causal inference in the context of observational studies, environmental epidemiology, and comparative effectiveness research. He received his Ph.D. from the Department of Biostatistics at the University of California, Los Angeles in 2010, and became faculty in the Harvard T.H. Chan School of Public Health in 2013.

**Chanmin Kim**, Ph.D., is a postdoctoral fellow in the Department of Biostatistics at the Harvard T.H. Chan School of Public Health specializing in the development of Bayesian nonparametric techniques and causal mediation analysis. He received his Ph.D. from the Department of Statistics at the University of Florida.

**Christine Choirat**, Ph.D., is a research scientist in the Department of Biostatistics at the Harvard T.H. Chan School

of Public Health and the Harvard Institute for Quantitative Social Science. Her specialty is data science, statistics, computer science, and reproducible research. She received her Ph.D. from Université Paris Dauphine in 2003.

**John Barrett Hansen** is a recent graduate of Harvard College who contributed to this report while completing an honors thesis.

**Yun Wang**, Ph.D., is a senior research scientist in the Department of Biostatistics at the Harvard T.H. Chan School of Public Health and leads management of Medicare data sources.

**Lauren Hund**, Ph.D., is a research scientist at Sandia National Labs in Albuquerque, New Mexico. She worked on this report while completing her Ph.D. and postdoctoral fellowship at the Harvard T.H. Chan School of Public Health.

**Jonathan Samet**, M.D., M.S., is a pulmonary physician and epidemiologist who is currently a professor and Flora L. Thornton Chair for the Department of Preventive Medicine at the Keck School of Medicine at the University of Southern California and the director of the University of Southern California Institute for Global Health.

**Gary King**, Ph.D., is the Albert J. Weatherhead III University Professor at Harvard University. He is based in the Department of Government and serves as director of the Institute for Quantitative Social Science. King develops and applies empirical methods in many areas of social science research, focusing on innovations that span the range from statistical theory to practical application to reproducibility.

**Francesca Dominici**, Ph.D., is a professor of biostatistics and senior associate dean for research at the Harvard T.H. Chan School of Public Health. Her research has focused on the development of statistical methods for the analysis of large observational data with the ultimate goal of addressing important questions in environmental health science, health-related impacts of climate change, and comparative effectiveness research.

---

**OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH**


---

Published work emanating from this grant funding appears in the following:

Dominici F, Cefalu M. 2015. Confounding adjustment and exposure prediction in environmental epidemiology: additional insights. (Letter to the Editor) *Epidemiology* 26(2):e28.

Dominici F, Wang Y, Correia A, Ezzati M, Pope CA III, Dockery DW. 2015. Chemical composition of fine particulate matter and life expectancy in 95 US counties between 2002 and 2007. *Epidemiology* 26(4):556–564.

Wang C, Parmigiani G, Dominici F, Zigler CM. 2015. Accounting for uncertainty in confounder and effect modifier selection when estimating average causal effects in generalized linear models. *Biometrics*; doi: 10.1111/biom.12315.

Zigler, CM. 2015. The central role of Bayes theorem for joint estimation of causal effects and propensity scores. *Am Stat*; doi: 10.1080/00031305.2015.1111260.

Cefalu M, Dominici F. 2014. Does exposure prediction bias health-effect estimation? The relationship between confounding adjustment and exposure prediction. *Epidemiology* 25:583–590.

Dominici F, Greenstone M, Sunstein CR. 2014. Science and regulation. Particulate matter matters. *Science* 344(6181): 257–259.

Zigler CM, Dominici F. 2014. Clarifying policy evidence with potential-outcomes thinking: beyond exposure–response estimation in air pollution epidemiology. *Am J Epidemiol* 180(12):1133–1140.

Zigler CM, Dominici F. 2014. Uncertainty in propensity score estimation: Bayesian methods for variable selection and model averaged causal effects. *J Am Stat Assoc* 109(505):95–107.

Zigler CM, Watts K, Yeh RW, Wang Y, Coull BA, Dominici F. 2013. Model feedback in Bayesian propensity score estimation. *Biometrics* 69(1):263–273.

---

**ABBREVIATIONS AND OTHER TERMS**


---

AMPD	Air Markets Program Data
AQS	air quality system
AREPA	a software package that retrieves and processes EPA data
ARP	Acid Rain Program
ATT	average treatment effect on the treated
CAA	(U.S.) Clean Air Act
CEM	continuous-emissions monitor
CM	clinical modification
CMS	Centers for Medicare and Medicaid Services
CO	carbon monoxide
CO <sub>2</sub>	carbon dioxide
CVD	cardiovascular disease
EAE	expected associative effects
EDE	expected dissociative effects
EGU	electricity-generating unit
EIA	(U.S.) Energy Information Administration
ICD-CM	International Classification of Diseases–Clinical Modification
JNIE	joint natural indirect effect
NAAQS	National Ambient Air Quality Standard
NDE	natural direct effects
NIE	natural indirect effects
NO <sub>2</sub>	nitrogen dioxide
NO <sub>x</sub>	nitrogen oxides
O <sub>3</sub>	ozone
PM <sub>10</sub>	particulate matter ≤ 10 μm in aerodynamic diameter
PM <sub>2.5</sub>	particulate matter ≤ 2.5 μm in aerodynamic diameter
SIP	state implementation plan
SO <sub>2</sub>	sulfur dioxide
TE	total effect
U.S. EPA	U.S. Environmental Protection Agency



Research Report 187, *Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations*, C.M. Zigler et al.

---

INTRODUCTION AND SCIENTIFIC BACKGROUND

---

The work of Dr. Corwin M. Zigler and colleagues, *Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations*, was funded under RFA-11-1, the latest in a series of requests for applications issued as part of HEI's Accountability research program (see Preface for a history of this program). Accountability research is designed to evaluate the extent to which air quality regulations have succeeded in improving air quality and/or public health.

Interest in assessing the health effects of air quality actions has grown in response to questions about the benefit of tightening air pollution regulations. Since the 1980s, measurements at thousands of monitoring stations across the country have shown reduced concentrations of all six criteria pollutants. This progress has come at a price. The U.S. Environmental Protection Agency (U.S. EPA\*) estimated that from 1970 to 1990 the cost of air pollution control was about \$25 billion per year. Even as new research has strengthened evidence for both adverse health effects of air pollution and the case for regulatory and other preventive measures, and even as estimates of health benefits have exceeded the estimated costs of regulatory interventions (Office of Management and Budget 2010), policymakers, legislators, industry representatives, and the U.S. EPA continually seek to document whether past efforts to reduce air pollution have yielded demonstrable improvements in public health and to better predict whether future efforts will continue to do so.

Assessments of the benefits of air quality regulations have generally relied on concentration–response functions

---

Dr. Corwin M. Zigler's 3-year study, "Causal Inference Methods for Estimating Long Term Health Effects of Air Quality Regulations," began in April 2012. Total expenditures were \$885,962. The draft Investigators' Report from Zigler and colleagues was received for review in May 2015. A revised report, received in September 2015, was accepted for publication in October 2015. 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 Critique. (As a coinvestigator of the Zigler report, Dr. Francesca Dominici was not involved in its selection by the Research Committee.)

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

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

from retrospective epidemiologic studies, which are used to predict public health outcomes that might occur under alternative air pollution control scenarios (U.S. EPA 1999). Accountability research has been developed to assess whether such estimates can be validated against the outcomes in real-world studies of actual regulatory programs and other interventions. In its first publication on accountability, the HEI Accountability Working Group (2003) set out a conceptual framework for assessing the health effects of air quality regulations that has since formed the basis for its research program. An important feature of that framework was the "chain of accountability," a simplified causal model of the interrelations between particular interventions and their effects on pollutant emissions, concentrations, exposures, and health outcomes.

Over the past 15 years, studies funded under HEI's Accountability research program and by others have primarily focused on effects of short-term interventions. Some studies have examined the impact of activities not necessarily focused on air quality (e.g., traffic control around the Atlanta Olympics [Peel et al. 2010] and the London Congestion Charging Scheme [Kelly et al. 2011]). Others examined the impact of short- and somewhat longer-term interventions intended specifically to improve air quality and health (e.g., respectively, Zhang et al. 2013 study of traffic and other industrial emissions controls instituted for the 2008 Beijing Olympics, Dockery et al. 2013 study of the extension of the coal bans, first studied in Dublin by Clancy et al. 2002 and Hedley et al. 2002, to additional Irish counties).

These studies encountered a number of challenges. Studies of interventions aimed at reducing traffic congestion have found only small improvements in air quality; reducing the statistical power with which to continue to a direct evaluation of health effects (Kelly et al. 2011). In other studies, researchers have found that the observed air quality changes were regional in nature and could not be directly related to the intervention at the local level (Peel et al. 2010). The results of studies of the health effects of air-quality-improvement programs implemented over short time frames (e.g., Clancy et al. 2002 and Hedley et al. 2002) suggest that the outcomes of such interventions may be directly measurable after a relatively short time period if a substantial change in air quality is produced. In the case of

the Beijing Olympics, the HEI Health Review Committee agreed that the changes in air quality observed during the Olympics were consistent with a successful intervention but that the study was not designed (e.g., with control communities) to assess whether the control measures could be considered causal either in producing the changes in pollutant levels or in the changes in biomarkers that had been measured. After adding control communities to his extended analysis of the coal bans in Ireland, Dockery and colleagues (2013) demonstrated the challenges of disentangling the effects of an intervention from those of other social and economic factors that might also influence long-term trends in air quality and health.

With RFA-11-1, HEI sought to extend its Accountability research program to large-scale, multiyear regulatory programs designed to improve air quality and health, in particular those aimed at large urban areas and major ports. The study of such long-term actions is particularly important because their estimated health benefits tend to dominate the overall estimated benefits of air pollution regulation. However, because of the considerable challenges inherent in such research, few studies have been undertaken to date (e.g., Auffhammer et al. 2009; Chay and Greenstone 2003; Harrington et al. 2010, 2012; Moore et al. 2010; Pope et al. 2009).

The challenges posed in evaluating long-term regulatory actions have been summarized in HEI Communications 14 (van Erp and Cohen 2009) and 15 (Health Effects Institute 2010) and demonstrated in the recent accountability studies discussed above. Regulatory interventions to improve air quality, especially large national programs such as the U.S. Clean Air Act, may not have immediate effects on either air quality or public health. Ensuing changes in emissions, ambient concentrations, and human exposure may occur at different times and on different spatial scales (e.g., local, state, and national), and the dynamics of biological processes of injury that underlie adverse health effects of air pollution may not directly follow the changes in exposure that result from regulatory action. The longer the time between promulgation of a regulation and its effects, the greater the possibility that other factors that influence health outcomes (e.g., changes in medical practices and the availability of health care) may come into play and interfere with demonstrating the effects of the regulation itself. The degree to which the regulation is enforced may further complicate the analysis by extending the anticipated time between intervention and effect. Therefore, it may be difficult to isolate the causal pathways leading from regulation and its consequences for air quality to a change in health risk.

One of the conclusions from HEI's assessments of the accountability literature was that further development of

suitable epidemiologic and statistical approaches would be necessary to support the evaluation of long-term regulatory actions. One of these approaches, implicit in HEI's advice to and comments on studies by Peel (2010), Zhang (2013), and Dockery (2013) and their colleagues was to consider designs including suitable controls that would enhance the ability to attribute the changes in air quality and health to the interventions under study.

In response to RFA 11-1, Dr. Zigler and colleagues submitted their proposal, "Causal Inference Methods for Estimating Long Term Health Effects of Air Quality Regulations," in 2011. The Research Committee was very enthusiastic about the proposal because the Committee thought that the methodological advances in causal inference would be substantial and interesting to not only the HEI community but to the broader statistical community. The Committee also liked the proposed demonstration of these methods in real-world interventions. Because of reservations the Committee had about the ambiguities associated with the first policy intervention proposed for study (i.e., nonattainment designation with particulate matter  $\leq 10 \mu\text{m}$  in aerodynamic diameter [ $\text{PM}_{10}$ ]), the Committee recommended that the research team add someone with expertise in the implementation of the U.S. Clean Air Act. In response, Dr. Zigler added as a consultant Mr. John Bachmann, former Associate Director for Science/Policy and New Programs for the U.S. EPA's Office of Air and Radiation. They also planned to work with Mr. Bachmann to identify other, less ambiguous actions for which observed data could support estimation of causal effects.

This Critique 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 (IR) into scientific and regulatory perspective. It begins with a brief overview of the study.

---

## SUMMARY OF THE STUDY

---

### SPECIFIC AIMS

The investigators outlined three specific aims:

1. Use a potential outcomes framework to define causal effects of interest for single-pollutant accountability assessment and develop methods for estimation.
2. Define causal effects for multipollutant accountability assessment and develop methods for estimation.
3. Develop national databases, conduct epidemiological studies, and disseminate software and results.

## METHODS

### Overview of the Causal Methods Development

The overall goal of the study was to develop what the investigators refer to as “direct accountability assessment” of the effectiveness of particular regulatory interventions. “Direct” methods stand in contradistinction to “indirect” methods, which rely on associations between exposure and outcome from epidemiologic studies to predict the benefits of a proposed intervention relative to a counterfactual situation in which no intervention might be implemented. The authors used both “established methods for drawing causal inferences from observational data and newly developed methodology for causal accountability assessment.”

In concert with other published work on causal methods, the first feature of their approach was the reframing of air pollution interventions as a hypothetical randomized experiment, analogous to a randomized clinical trial in which some subjects are randomly assigned to receive “treatment” and others to receive no treatment (i.e., the “controls”). Randomized studies are considered the optimal study design for determining the efficacy, or causal influence, of treatment because randomization typically results in similarity between the treatment and control groups with respect to confounding factors that might also affect disease outcome. As the HEI Accountability Working Group noted in 2003, randomized allocation of reductions in air pollution concentrations has neither been feasible or ethical in the United States. Consequently, in the “direct accountability assessment” undertaken by Zigler and his colleagues, the authors “approximate” this randomized design by comparing portions of the population experiencing the intervention (the treatment) to other portions of the population that did not receive the intervention (the controls) but are otherwise comparable on a number of factors or characteristics that have been measured. In the “indirect” accountability approach typically used in benefit–cost analysis, the future health benefits of an intervention are estimated from the intervention’s projected impact on exposures combined with the exposure–response functions derived from retrospective epidemiological studies.

The second feature of their approach was their analysis of causal pathways for each intervention. The causal pathways are the pathways through which an intervention may act to cause changes in the outcome of interest, whether a reduction in emissions, air pollution, or health outcomes. The pathway may represent the direct effect of one factor on an outcome (e.g., air pollution on health outcomes) or may involve the mediation of the effect of that factor through some intermediate step or factor. The investigators

used two different but conceptually analogous methods, principal stratification and causal mediation analyses to investigate causal pathways.

Principal stratification involves comparison of outcomes between key strata or groupings of the data (for example, the effects on health in areas where an intervention has caused a reduction in air pollution and those where it has not). Using this general example, it defines “associative” effects as those effects on health that occur when an intervention caused a meaningful reduction in air pollution and “dissociative” effects as the effects on health outcomes that occur when the intervention did not have a causal effect on air pollution. The size of the associative effects relative to the dissociative effects provides an indication of the relative importance of the two pathways, in this example an indication of the intermediate role of the reduction in air pollution. Causal mediation methods are also designed to evaluate the effect of mediators or intermediate steps on an outcome of interest but in a more formal way. Using our general air pollution example, causal mediation divides the effects of an intervention into two components: (1) the “natural direct effect,” defined as the direct effect of the intervention on the outcome, and (2) the “natural indirect effect,” defined as the causal effect mediated by changes in some intermediate factor like a specific air pollutant. However, unlike in principal stratification, these two effects sum to the total effect.

The authors demonstrated the use of these methods (see Critique Table 1) in two case studies designed to assess the outcomes of different regulatory interventions (Case Study 1: Accountability Assessment of PM<sub>10</sub> Nonattainment Designations in the Western United States and Case Study 2: Accountability Assessment of Power-Plant Emissions Controls). Each case study addressed somewhat different steps in the “chain of accountability”; neither study addressed all of the steps (see Critique Table 2). The Committee’s summaries of the methods and findings for each study are found below followed by the Committee’s Critique of the report.

### CASE STUDY 1: ACCOUNTABILITY ASSESSMENT OF PM<sub>10</sub> NONATTAINMENT DESIGNATIONS IN THE WESTERN UNITED STATES

#### Introduction

In the first case study, the authors evaluated the effect on air quality and on health outcomes of designating areas of the Western United States to be in “nonattainment” in the period 1990–1995 with the 1987 National Ambient Air Quality Standard for PM<sub>10</sub>. The intervention in this case — designating counties to be in nonattainment — required the state to develop a state implementation plan that

**Critique Table 1.** Overview of Causal Inference Methods Used in Case Study 1 and Case Study 2

Causal Inference Methods	Case Study 1	Case Study 2
Potential outcomes framework (Framing as a randomized experiment)	X	X
Propensity scores (confounding)	X	
Principal stratification	X	X (multipollutant)
Causal mediation analysis		X (multipollutant)

**Critique Table 2.** Links of the Chain of Accountability Addressed by Case Study 1 and Case Study 2

Successive Links in the Chain of Accountability	Case Study 1	Case Study 2
Regulatory action	X	X
Emissions		X
Ambient air quality	X	X
Exposures / doses		
Human health responses	X	

outlined the strategies to be taken to bring the counties into compliance by 2001. This case study involved only three of the links in the “chain of accountability”: an intervention, change in air quality, and a human health response (see Critique Table 2); it did not estimate the intermediate impacts on emissions and on estimated exposures.

The authors obtained air quality data for the years 1990–2001 from 547 monitors in the U.S. EPA Air Quality System database; 268 monitors were located in nonattainment areas. They assessed impact on air quality in terms of annual average changes in PM<sub>10</sub> concentrations from 1999–2001. The authors used the Medicare Part A hospital claims and enrollment data to assemble information on demographic, all-cause mortality, and hospitalization information, in particular hospitalizations for cardiovascular and respiratory-related illnesses using specific International Classification of Diseases–9 (ICD-9) codes.

To conduct their analyses, the authors created a national linked database comprising data on the population of 3,971,610 U.S. Medicare beneficiaries living within 6 miles of a PM<sub>10</sub> monitoring station in 2001; 2,349,691 beneficiaries, about 60% of the total, lived in nonattainment areas. Medicare beneficiaries and their data were linked to data

from the monitors nearest to their zip code of residence. The database also included county-level demographics, urbanicity, and smoking data from the 2000 U.S. Census, from the Centers for Disease Control and Prevention’s Behavioral Risk Factor Surveillance Survey, and county-level temperature data from the National Climactic Data Center.

### Statistical Methods

The authors defined the primary causal question for this case study to be “Are Medicare health outcomes in PM<sub>10</sub> nonattainment areas different from what they would have been if the nonattainment designations had never occurred?” In the framing of the analysis like a randomized controlled experiment, areas that received nonattainment designations were considered to be assigned to “treatment” whereas areas that were in attainment were assumed to be “controls.” Because these two groups were in fact not selected via a randomized process, the authors chose to develop and use propensity scores to adjust for possible differences between the two groups that could confound the relationship. Propensity scores are an aggregate measure of multiple underlying covariates (Robins et al. 2000; Rosenbaum and Rubin 1983; Rubin 2008; Stuart 2010), that the authors identified and assumed to include or to represent

all possible confounding factors. Thus, when propensity scores are similar for the “treatment” and “control” group, it is assumed that little potential for confounding exists. The authors “pruned” from the analysis 52 areas (of the original 547 areas) where the propensity scores were not comparable to the opposite intervention group. The remaining 495 areas (276 attainment, 219 nonattainment) were grouped by comparable propensity scores into five subclasses. They used a spatial hierarchical regression model for log transformed  $PM_{10}$  and log-linear Poisson regression models for mortality and hospitalization outcomes to predict what the  $PM_{10}$  concentrations and the health outcomes, respectively, might have been in the 219 nonattainment areas had no designation been made — that is, the overall causal effects of designation as a nonattainment. The health endpoints evaluated included all-cause mortality and hospitalizations for cardiovascular disease and respiratory diseases. All models were adjusted for the propensity scores subclasses discussed above and for selected additional covariates. Markov chain Monte Carlo was used for estimation. These methods were described in Appendix A to the report (available on the HEI Web site) and in a published paper (Zigler et al. 2012).

Their second objective was to evaluate whether the effects on health outcomes of designation as a nonattainment area were influenced by reduction in ambient  $PM_{10}$  concentrations or by other unidentified factors. To evaluate the role of  $PM_{10}$ , they used principal stratification analysis in which they defined “associative” effects as the effects on health when the nonattainment designation was found to cause a reduction by at least  $5 \mu\text{g}/\text{m}^3$  of ambient  $PM_{10}$ , and “dissociative” effects as the effects on health outcomes that occurred when the designation did not have a causal effect on  $PM_{10}$ .

The greater the associative effect compared with the dissociative effect, the more  $PM_{10}$  can be inferred to have a role in the causal pathway leading to improvement of health outcomes. Results are presented in terms of the posterior mean and 95% posterior intervals.

## Results

Zigler and colleagues estimated that nonattainment designation had causally reduced  $PM_{10}$  concentrations by an average (posterior mean) of  $1.17 \mu\text{g}/\text{m}^3$  (95% posterior interval:  $-7.33, 4.00$ ) for the period 1999–2001. They concluded that there was evidence of a causal reduction in all-cause mortality of 1.08 deaths per 1000 Medicare beneficiaries ( $-1.08; -3.27, 0.99$ ) and in respiratory hospitalizations ( $-1.47; -3.86, 0.70$ ) per 1000 person-years. However, they concluded there was no evidence of a causal effect on cardiovascular hospitalizations ( $1.44; -4.64, 7.16$ ) per 1000 person-years.

With their principal stratification analysis, presented in IR Figure 9, Zigler and colleagues found differing results for the intermediary role of  $PM_{10}$  in causal effects on the three health outcomes they examined. For all-cause mortality, the dissociative effect ( $-1.90$  deaths per 1000 beneficiaries:  $-5.52, 1.87$ ) was on average slightly greater than that of the associative effect ( $-0.46; -4.03, 2.64$ ). This pattern suggests some reduction in mortality even in areas where  $PM_{10}$  was not causally affected. For respiratory outcomes, however, they found associative effects greater than dissociative effects in areas where  $PM_{10}$  had been causally affected, suggesting that  $PM_{10}$  played a causal role in the reduction of hospitalization for respiratory disease. In that case, the posterior mean associative effect was  $-3.34$  hospitalizations per 1000 person-years ( $-7.43, 0.67$ ) compared with  $-0.31$  ( $-3.84, 3.18$ ) for the dissociative effect. The dissociative and associative effects for cardiovascular hospitalizations were  $2.83$  ( $-5.84, 11.01$ ) and  $-3.78$  ( $-11.69, 3.79$ ) hospitalizations per 1000 person-years, respectively. A sensitivity analysis using  $1.2 \mu\text{g}/\text{m}^3$  as a cut point for a substantial effect of nonattainment designation (the average  $PM_{10}$  reduction detected in the data, rather than the  $5 \mu\text{g}/\text{m}^3$  first chosen by the authors), resulted in a similar pattern of dissociative and associative effects but with greater uncertainty (broader 95% posterior intervals).

The authors suggested that the observed causal effect of nonattainment designation on mortality, in the absence of a strong associative effect for  $PM_{10}$ , may be attributable to causal pathways other than the one involving reduction of  $PM_{10}$ . However, they noted several factors that contributed to uncertainty in their results, in particular the challenges in defining a truly definitive or discrete intervention:

1. Designation as nonattainment is akin to the “intention to treat” concept in clinical trials. Actions may or may not occur and may occur on different time frames.
2. The types of actions taken may differ between nonattainment areas.
3. Air quality control can have regional impacts such that nonattainment and attainment areas are both affected.

## CASE STUDY 2: ACCOUNTABILITY ASSESSMENT OF POWER-PLANT EMISSIONS CONTROLS

### Introduction

The second case study was designed to evaluate the causal impacts of installing a range of scrubber technologies on coal-fired power plants pursuant to requirements to reduce sulfur dioxide ( $\text{SO}_2$ ), nitrogen oxides ( $\text{NO}_x$ ), and carbon dioxide ( $\text{CO}_2$ ) emissions under the Acid Rain

Program, a program created by the 1990 amendments to the Clean Air Act. As the authors discuss, numerous studies have projected large health benefits from the Acid Rain Program because of projected reductions in power plant-related particulate matter  $\leq 2.5 \mu\text{m}$  in aerodynamic diameter ( $\text{PM}_{2.5}$ ). As a clearly defined intervention, whose effects on reducing  $\text{SO}_2$ ,  $\text{NO}_x$ , and  $\text{PM}_{2.5}$  have been well studied, scrubbers provide a good opportunity for a “proof of concept,” for new analytical methods. This case study thus addressed three links in the chain of accountability (see Critique Table 2); the intervention, the emissions, and ambient air quality. Efforts to extend these analyses to the evaluation of health outcomes are still ongoing.

The authors obtained annual emissions data for the year 2005 from continuous emissions monitors at 258 power plants. They also collected data on other potentially relevant covariates for the analysis: information on temperature and barometric pressure (from 2004); total plant emissions of  $\text{SO}_2$ ,  $\text{NO}_x$  and  $\text{CO}_2$ ; the use of  $\text{NO}_x$  emissions controls; coal sulfur content; and total operating time of individual energy-generating units within each plant in 2004 (see IR Table 4). They linked each power plant to all ambient  $\text{PM}_{2.5}$  monitors located within 150 km, although no monitor was assigned to more than one plant. Data from these monitors were used to calculate the average annual ambient  $\text{PM}_{2.5}$  concentration in 2005 for each set of monitors linked to a particular plant. All of these data were assembled and linked for analysis within a national database.

### Statistical Methods

For purposes of the intervention, a facility was considered to be “treated” if at least 10% of its total heat input came from energy-generating units within the facility that had been equipped with scrubbers as of January 2005. Those facilities with less than 10% of the total heat input from energy-generating units equipped with scrubbers served as “controls.” Of the 258 power plants for which annual emissions data were obtained, 63 met the definition of “treated,” 195 did not. In the potential outcomes framework, the causal effect of scrubber installation was defined as a comparison of emissions of  $\text{SO}_2$ ,  $\text{NO}_x$ , and  $\text{CO}_2$  observed with (“treatment”) and without (“control”) the installation of a scrubber.

Their multipollutant causal pathway analysis was structured to answer the question, “To what extent is the causal effect of a scrubber (the “treatment”) on ambient  $\text{PM}_{2.5}$  (the “outcome”) mediated through reduced emissions of  $\text{SO}_2$ ,  $\text{NO}_x$ , and  $\text{CO}_2$  (the “mediators”)?” Zigler and colleagues applied both principal stratification and causal mediation methods separately to answer this question. Given what is known about scrubbers, their expectation was that their effects on  $\text{PM}_{2.5}$  would be mediated primarily by  $\text{SO}_2$ .

The principal stratification analysis defines effects based only on the observable outcomes — the emissions and  $\text{PM}_{2.5}$  concentrations related to power plants with and without scrubbers. As in the first case study, the causal effects of the scrubber on  $\text{PM}_{2.5}$  were analyzed in terms of associative effects, the causal effects of a scrubber on ambient  $\text{PM}_{2.5}$  among power plants where emissions of  $\text{SO}_2$ ,  $\text{NO}_x$ , and  $\text{CO}_2$  were causally or “meaningfully affected” by the presence of a scrubber, and by dissociative effects, the causal effects on  $\text{PM}_{2.5}$  where emissions were not “meaningfully affected.” They chose as a cut point between “meaningfully” and “not meaningfully affected” to be half of the standard deviation of the estimated individual-level causal effect of a scrubber on the emission the  $k$ th pollutant ( $0.5 \sigma_k$ ).

The causal mediation analysis estimated the “natural direct” and “natural indirect” effects of the installation of a scrubber. The natural direct effect is defined as the effect on  $\text{PM}_{2.5}$  by the scrubber directly (i.e., not attributable to changes in emissions of the  $\text{SO}_2$ ,  $\text{NO}_x$ , and  $\text{CO}_2$ ), whereas the natural indirect effects are defined as those causal effects mediated only by changes in the emissions of the three pollutants, either individually or in various combinations with each other. Zigler and colleagues developed new Bayesian nonparametric methods for estimation of the natural direct and indirect effects under a set of key assumptions listed in IR Table 5. Although the direct and indirect effects are conceptually similar to the dissociative and associative effects in the principal stratification analysis, a key difference is that in the causal mediation analysis, the effect is based on a comparison between outcomes that are observed and those that are “unobservable,” (also referred to as a priori counterfactuals) and cannot be observed in reality. For example, one “unobservable outcome” would be the level of ambient  $\text{PM}_{2.5}$  expected in the scenario where a scrubber is installed but  $\text{SO}_2$  emissions are assumed to remain at the level they would have been had scrubbers not been installed.

### Results

Using their potential accountability approach, Zigler and colleagues estimated that scrubbers had causal effects on  $\text{SO}_2$ , but not on  $\text{NO}_x$  and  $\text{CO}_2$ . The posterior mean  $\text{SO}_2$  emissions were  $-1.02$  log tons lower ( $-1.41, -0.68$ ) from power plants with scrubbers than those without scrubbers, whereas the emissions of  $\text{NO}_x$  and  $\text{CO}_2$ ,  $0.12$  log tons ( $-0.15, 0.37$ ) and  $0.103$  log tons ( $-0.09, 0.29$ ), respectively, appeared to be unaffected by installation of scrubbers. They also argued that scrubbers had a causal effect on ambient  $\text{PM}_{2.5}$  concentrations (the total effect); on average scrubbers reduced  $\text{PM}_{2.5}$  concentrations relative to what their concentrations might have been absent the scrubbers ( $-0.38 \mu\text{g}/\text{m}^3: -1.25, 0.48$ ).

The multipollutant extension of principal stratification analysis, which estimated the associative and dissociative effects of SO<sub>2</sub> scrubbers on PM<sub>2.5</sub> mediated by all three pollutants, provided similar yet more complex insights (the results are detailed in IR Table 6 and IR Figure 13). In general, the authors concluded from their results that the installation of scrubbers causally reduced ambient PM<sub>2.5</sub>, on average, principally by means of SO<sub>2</sub> reduction (posterior mean associative effect  $-0.596 \mu\text{g}/\text{m}^3$ ;  $-1.1643, 0.295$ ). They argued that their results provide support for the conclusion that causal reductions of ambient PM<sub>2.5</sub> are not mediated by reductions in NO<sub>x</sub> and CO<sub>2</sub> (i.e., the dissociative effects were similar to or greater than the associative effects). Graphical analyses using three-dimensional surface plots also suggested that the causal effects on PM<sub>2.5</sub> were greatest where reductions in SO<sub>2</sub> were the greatest (IR Figure 14). Scrubbers were estimated to cause an increase in ambient PM<sub>2.5</sub> for a small proportion of plants for which SO<sub>2</sub> emissions had actually declined. The authors noted, however, that their conclusions were tempered by the fact that 95% posterior intervals for all the results in the principal stratification analysis were quite broad and included zero (see IR Table 6).

The results of the causal mediation analysis (IR Table 7) were broadly similar to those of the principal stratification analysis. Zigler and colleagues concluded that the causal effect on ambient PM<sub>2.5</sub> concentrations (the total effect) was primarily mediated through reduction in SO<sub>2</sub> because the natural indirect effect based on SO<sub>2</sub> had a posterior mean of  $-0.408 \mu\text{g}/\text{m}^3$  (95% posterior interval:  $-0.589, -0.262$ ), whereas the natural indirect effects mediated by NO<sub>x</sub> and CO<sub>2</sub> and the natural direct effects (without mediation by any pollutant) were close to zero. Similarly, the joint natural indirect effects of two or more of the pollutants were comparable to that of SO<sub>2</sub> alone only when SO<sub>2</sub> was one of the pollutants. The joint natural indirect effects were otherwise near zero.

Overall, the investigators concluded that their findings for this case study are internally consistent as well as consistent with the external scientific evidence on the role of scrubbers in reducing SO<sub>2</sub> emissions and ambient PM<sub>2.5</sub> concentrations. They argue that the case study was a reasonable demonstration of their newly developed causal methods for multiple, nonindependent pollutants.

---

## HEALTH REVIEW COMMITTEE'S CRITIQUE

---

With this study, Zigler and his colleagues tackled several important issues. In accordance with the goals of HEI's Accountability research program and RFA-11-1, they set out

to evaluate the outcomes of long-term regulatory actions. Unlike the other projects funded under this RFA, they had serious statistical methodological objectives for their accountability research, seeking to ground it in a potential outcomes framework, to apply new causal inference methods, and to extend those methods to multipollutants and to human health. The HEI Health Review Committee reflected first on the methodological contributions and then on the implications for accountability research on long-term regulatory actions.

In the Committee's view, the major contribution of this report is that it provides a very well-written primer on the application of a potential accountability framework and causal inference methods to the study of air pollution interventions. The investigators formally reframed the way we might think about assessing the outcomes of intervention to improve air pollution from the "indirect" approach currently used in regulatory benefits assessment to a "direct" approach, using the concepts from randomized studies and methods from causal inference. They extended air pollution "accountability assessment" by providing an explicit way to evaluate relationships between an action and associated outcomes that inherently considers what outcomes might have occurred in the absence of the action. In doing so, they avoid some of the pitfalls of earlier accountability studies that have simply observed associations.

Finally, they provide an excellent discussion (see "A Note on the Word 'Causal'") of what "causal" means — and does not mean — in the context of their statistical methods and its relationship to broader scientific evidence used in the evaluation of causal relationships. The Committee agrees with the authors that it is important to be clear about the limits of what "causal" means in the context of their statistical methods development. The investigators noted that their "estimates are not automatically guaranteed to have causal validity; rather, our framework provides a rigorous and principled way of clarifying and remedying some of the most common threats to validity that have plagued epidemiological studies... . As in any epidemiological study, the estimates should be interpreted in light of the available data and the specifics of the statistical models used for estimation."

The underlying causal analysis methods — propensity scores, principal stratification, causal mediation — that the investigators brought to their analysis of interventions under this framework are not new. However, the Committee thought it important to recognize that the application of these methods to air pollution accountability research was new and a non-trivial accomplishment to apply on the large scale of the two interventions they evaluated. The investigators needed to develop the software specifically for

each of the case studies and for working with the underlying data available to support them. The reason these methods have not been widely applied before may be not just the lack of knowledge of their existence but the operational impediments to applying them. The investigators should be commended for their plan to make publicly available the statistical code necessary both to assemble and link their data sources and to implement their newly developed methods.

Zigler and colleagues followed a standard but rigorous process for the development of their causal inference approach. They applied the methods in two well-developed case studies of regulatory interventions — the designation of nonattainment for PM<sub>10</sub> National Ambient Air Quality Standards in case study 1 and the installation of scrubbers under the Acid Rain Program in case study 2 (see Critique Table 1 for an overview of methods). Their careful application of the methods to two well-developed case studies illustrates that their approaches are feasible. However, as the Committee's discussion of the individual studies shows, the investigators have also identified several important challenges to implementing their approaches in complex real-world interventions that hold lessons for future research.

#### **CASE STUDY 1: IMPACTS OF DESIGNATION AS NONATTAINMENT AREA ON PM<sub>10</sub> AND HEALTH**

The first case study is the largest effort to date to examine carefully the impact of designation of areas to be in nonattainment with national ambient air quality and the first to examine this action in a potential outcomes framework. Zigler and his colleagues examined the impact on nonattainment in a large subsection of the United States, the Western United States. In their analysis, they addressed three links in the chain of accountability from the initial “intervention” or “regulatory action” (nonattainment designation), to impact on ambient air quality (specifically PM<sub>10</sub> concentrations), to human health responses (see Critique Table 2).

The use of the Medicare database was useful and appropriate for this application because it is a national database and geographically representative. The database also includes those individuals in the population who are more likely to be susceptible to the adverse health outcomes whose relationship to air pollution were being studied (i.e., all-cause mortality and hospitalizations from cardiovascular or respiratory disease).

The application of statistical methods in the case study was logical and sound. Much of the technical detail for the first case study is provided in detailed appendices to the report (Appendices A and B, available online) and in a

series of published papers (see IR section Other Publications Resulting from this Research). This approach to reporting the details of the first case study had the advantage of making the report more readable for a general audience but more difficult to ascertain some of the key details and assumptions of the methods.

The investigators' use of propensity scores to deal with possible confounding was appropriate. Different characteristics of the attainment and nonattainment areas were reflected in differences in the estimated propensity scores, which became the basis for removing non-comparable monitoring locations from the analysis. These monitoring locations would not have been removed in a normal regression analysis. At the Committee's request, the investigators also added a useful sensitivity analysis where these locations were not removed from the analysis.

However, one of the key challenges the investigators faced is that they had to make what they acknowledge to be a strong assumption — that they had identified all important confounders. That is, their propensity score method assumed that the covariates identified in IR Table 1 constituted, or were proxies for, all factors that could confound comparisons between attainment and nonattainment areas. However, there are reasons to question this assumption. For example, current or prior designation of nonattainment of another pollutant (e.g., ozone or total suspended particulates, or violation of a 24-hour standard for PM<sub>10</sub>) could be quite important as the sources targeted to reduce these pollutants can also impact annual average PM<sub>10</sub>. The Committee raised the question as to whether temporal within-unit comparisons would be less affected by confounding (although temporal trends in factors affecting the outcomes would have to be addressed). Furthermore, taking the outcome to be the 1999–2001 average PM<sub>10</sub> concentration, rather than the whole time series of PM<sub>10</sub> measurements did not utilize all the available outcome data, which may have reduced the power of the analysis.

The authors conducted a relatively straightforward application of principal stratification analysis to identify the causal pathways through which nonattainment designations might influence health outcomes. In this case, they examined two pathways, one that operated via causal reductions in annual average PM<sub>10</sub> levels (associative effects) and one that operated via other unknown mechanisms (dissociative effects). The relative magnitude of the associative and dissociative effects was interpreted as insight into the relative importance of air pollution compared with other factors in the observed impacts on health. However, a key challenge in this analysis was that the determination of what was a “causal reduction” in PM<sub>10</sub> concentrations was a somewhat arbitrary decision; the investigators initially

chose as the cut point a reduction of 5  $\mu\text{g}/\text{m}^3$  in  $\text{PM}_{10}$  concentrations although the average effect of nonattainment designation was a reduction of about 1.2  $\mu\text{g}/\text{m}^3$ . They ultimately added a useful sensitivity analysis to characterize the impact of choosing the actual effect on  $\text{PM}_{10}$  cut point, although other cut points might also have been examined.

Ultimately, perhaps the greatest challenges to demonstrating the use of the potential outcomes framework and causal inference methods in this case study were the choice and characterization of the “intervention” itself. As the authors pointed out early in their introduction, the goal of this work was to develop methods that would examine the effectiveness of “specific air quality regulations.” Designation of an area to be in nonattainment between 1990 and 1995, although having the advantage that it could be examined on a large spatial scale within the United States, is not an intervention in a classical sense. As the authors themselves noted, it is an ambiguous “intervention” with a number of sources of heterogeneity over space and time. A number of the most effective actions to address nonattainment in any one region are normally taken at the national rather than local level (e.g., vehicle emission control standards), and these may result in air quality improvements in both the nonattainment areas and the attainment areas. Some of the nonattainment areas were likely already implementing measures to improve air quality before the actual nonattainment designation (due to earlier nonattainment designations in 1987) and/or nonattainment designation for other pollutants besides  $\text{PM}_{10}$  (i.e., ozone). Indeed, the measurement data suggest improvements in air quality were already occurring prior to any implementation resulting from the 1990 nonattainment designation. Consequently, the results may reflect sensitivity to the authors’ decision to represent changes in  $\text{PM}_{10}$  by an aggregate measurement over a relatively short time period (i.e., the average annual concentration across 3 years encompassing 1999 to 2001).

The Committee thought that these challenges related to the choice of intervention do not reflect on the quality of work done to demonstrate the statistical concepts and methods but that they do contribute to limitations on the substantive conclusions that can be drawn from this case study as well as the general challenges in applying these methods to assess accountability. The Committee agrees with the authors that the results reflect the phenomenon of “signal modulation,” where heterogeneity in the intervention gets carried through and even amplified in heterogeneity of the estimated causal effects on air pollution and, in turn, on the effects on health outcomes. It is one explanation for the large 95% posterior intervals observed in the results. The Committee points out that the impact of this

phenomenon on uncertainty is substantial even when just three of the links in the chain of accountability are investigated. Inclusion of analyses on the other links could further add to the uncertainties.

The Committee agreed with the authors that the average effects reported suggest a reduction in  $\text{PM}_{10}$  concentrations as well as lower mortality rates and fewer hospitalizations for respiratory disease (but, surprisingly, an increase in hospitalizations from cardiovascular disease). However, given the degree of uncertainty in the findings, where the 95% posterior intervals included zero in all cases, the investigators have generally overstated the average causal effects of nonattainment designation on both  $\text{PM}_{10}$  concentrations and on each of the health outcomes.

The results of the principal stratification analyses also provided a mixed assessment of the causal role of  $\text{PM}_{10}$  reductions on the causal reductions in mortality and in hospitalizations (IR Figure 9). For all-cause mortality there was an overall reduction in mortality but the associative and dissociative effects were similar, indicating mortality was reduced in nonattainment areas but that this reduction was not clearly attributable to reductions in  $\text{PM}_{10}$  concentrations. For respiratory hospitalizations, which were on average lower, associative effects were larger than the dissociative effects, suggesting that the overall reduction was more attributable to  $\text{PM}_{10}$  reductions. Associative effects were also larger than the dissociative effects for cardiovascular hospitalizations, suggesting that the small increase in cardiovascular hospitalizations was causally related to decreases in  $\text{PM}_{10}$ . However, the Committee notes that because all these estimated effects include zero this analysis does not support strong conclusions in any direction.

In another sense, these broad uncertainty levels may more completely reflect the level of uncertainty in the estimated causal effects of an intervention than do methods that rely only on “indirect” accountability assessment. The principal stratification analysis in this case study, for example, allows for consideration of the evidence that a county may be designated to be in nonattainment, but that air pollution and mortality might not go down. In contrast, an indirect approach would assume that if a county were designated to be in nonattainment, it would receive the “treatment,” pollution would go down, and fewer deaths would ultimately result. Reality is more complex.

## CASE STUDY 2: SCRUBBER INSTALLATION ON COAL-FIRED POWER PLANTS

The scrubber case study provided a solid basis for development and testing of a richer set of statistical methods, Bayesian nonparametric principal stratification and causal

mediation applied in a multipollutant setting. The Committee thought the investigators' advances in statistical methodology were most clearly demonstrated here. In general, new statistical methods need to be tested in a setting that is reasonably well understood, such that study findings that do not comport with scientific knowledge signal some problem with the method or its application, including the underlying data available. The scientific principles underlying the efficacy of scrubbers for reducing emissions from power plants are well characterized and the impact of scrubbers on both emissions and on ambient particulate matter mass has been demonstrated.

An important contribution of case study 2 is that it provides a rare comparison of two different causal methods to the same real-world problem. In the words of one reviewer, "This is the first time I've seen such a comprehensive comparison of the principal stratification framework and the causal mediation approach to a non-trivial complex problem." While it was useful to see if their results agreed, it is also important to understand that these methods differ in important ways. A key difference is that principal stratification assumes that there is only one intervention (i.e., the scrubber), whereas causal mediation assumes there are two interventions (i.e., installation of the scrubber and the reduction of SO<sub>2</sub> levels). Another difference is that in principal stratification the total effect is decomposed into associative and dissociative effects, which might occur in different subgroups (e.g., groups of counties or power-generating units), whereas in causal mediation the total effect is decomposed into direct and indirect effects that apply to all groups in the study. Hence, direct and indirect effects are additive and the analysis could provide insights about the proportion of total effects that were attributable to one or the other.

The Committee thought the extension of the principal stratification and causal mediation analyses to multiple pollutants — that is, that the investigators examined whether scrubber-related reductions in PM<sub>2.5</sub> were mediated not only by SO<sub>2</sub> but also by CO<sub>2</sub> and NO<sub>x</sub> — was another important advance in the methods. Again, given the well-established scientific understanding of the role and impact of SO<sub>2</sub> scrubbers, this case study was appropriate to test the methodologies against expectations.

The demonstrations of these methodologies in the case study were transparent, well-conducted, and thorough. The investigators' decision to assemble and summarize the key assumptions required for estimation of causal effects by each of the methodologies in one table (IR Table 5) was a very useful addition to the revised report. The assumptions are standard for this area of statistical research. Whether they are reasonable is challenging, particularly

for the causal mediation framework, because they cannot really be verified with the observed data. However, having been presented clearly, their implications can be more directly debated. The Committee suggested that such a summary would be a welcome addition to any study of this kind.

The investigators also tested the sensitivity of their results to other specific analytical choices in their applications, for example, to linking power plants to all monitors within 75 km rather than 150 km, which was the choice made in the main analysis (see Appendix D, available online). In this case, they reported that the results were similar regardless of linkage distance, which adds some support to the assumption that the arbitrary choice of distance was not important. They noted that several other simplifying assumptions were made for the purposes of demonstrating the methods that might be explored more fully in future analyses. These include assumptions leading to exclusion of power plants within their study areas, as well as assumptions about the even distribution of impacts on PM<sub>2.5</sub> from emissions changes at these excluded power plants across areas in the study, the designation of plants as "treated" based on presence of a scrubber at a single point in time, the key attributes of the power plants that predict emissions, and the roles of topography, meteorology, and atmospheric chemistry. Given that the purpose of this study was to develop and demonstrate methods, the use of simplified examples was warranted, but may also have contributed to some of the uncertainties observed in the findings.

The Committee agreed with the investigators that their findings for this case study were broadly consistent with what would be expected given what is known about the role of scrubbers and their impact on SO<sub>2</sub> and ambient particulate matter. They found that installation of scrubbers on average caused reductions in SO<sub>2</sub>, rather than in CO<sub>2</sub> and NO<sub>x</sub>, and in ambient PM<sub>2.5</sub>. To varying degrees both principal stratification and causal mediation methods suggested that reductions in PM<sub>2.5</sub> were mediated through reductions in SO<sub>2</sub> and not through reductions in CO<sub>2</sub> or NO<sub>x</sub>, either individually or collectively. These methods technically estimate different parameters and have different interpretations (VanderWeele 2008). However, it was reassuring they both lead to similar conclusions when applied to the same data sets where the underlying mechanism is well understood.

At the same time, the Committee points out that the effects estimated by both methods in this case study are accompanied by a substantial degree of uncertainty. The 95% posterior intervals are very broad; as the investigators noted, the "causal" effect of scrubbers is "on average" a

reduction on  $PM_{2.5}$  but the effect is “not significantly different from zero.” At one level, the inability to find clearer or more significant signals reflecting relationships between  $SO_2$  scrubbers,  $SO_2$  emissions, and ambient  $PM_{2.5}$  is troubling. Both causal inference methods and Bayesian methods reflect more of the underlying uncertainties from each step of the analysis than do conventional confidence intervals. In that respect, they may provide a more realistic picture of the level of confidence with which the causal effects in this case can be predicted. The high degree of uncertainty may also be an indication of imprecision or inaccuracy in the investigators’ specification of the problem and the quality of the data available to conduct the analyses. We know, for example, that at an earlier stage in the development of the scrubber case study, the findings ran counter to expectations, suggesting that  $SO_2$  scrubbers effects on ambient  $PM_{2.5}$  were not causally mediated by changes in  $SO_2$  emissions. Re-examination of the scrubber assumptions with John Bachmann, a consultant with substantial expertise on scrubber applications, led to the addition of more covariates to characterize the scrubber units (e.g., use of selective non-catalytic reduction and the percent operating capacity). These changes helped lead to the findings that comported more closely with available evidence on the impacts of scrubbers on  $SO_2$  and  $PM_{2.5}$ .

This experience with what, it seems, should have been a relatively straightforward case study holds important lessons for the conduct of future studies, particularly for studies of problems in which the causal relationships are hypothesized but not fully known. One lesson is that it is extremely important to bring together collaborators with the necessary subject-area expertise to define the problem as clearly and accurately as possible. Another lesson is that it is equally important to understand and acknowledge the limitations of the underlying methods and data in the interpretation of the results. As we have seen in this case study, weak evidence of a causal effect is not necessarily strong evidence that such an effect does not exist.

The authors’ plans to extend this case study to the evaluation of the causal effects of scrubber installation on personal exposures to  $PM_{2.5}$  and on health outcomes (i.e., to the full chain of accountability) will likely encounter similar challenges. These analyses had been planned in the original proposal for this study but given the substantial challenges in getting the methods to work in the first three steps, the authors were unable complete this work within the time frame for this report.

---

## SUMMARY AND CONCLUSIONS

---

The Committee concluded that Zigler and his colleagues provided a well-conducted study and a well-written report on a difficult topic that makes a major contribution to the field of accountability research in the context of air pollution and health. The formal causal framework described in this report provides a particularly clear way forward for thinking about the health impacts of all kinds of interventions designed to reduce emissions and ambient air pollution, not just the “natural experiments” that have often been the focus of accountability studies in the past.

In addition to the formal causal framework, the statistical methods applied by the authors provide more information that can be used to think about the problem and to explore alternative causal pathways than do the “indirect” assessments in epidemiology studies, which have not been obtained under a causal framework. Another feature of the methods is that, compared with “indirect” assessments, “direct” approaches will include a more complete characterization of uncertainty. In part, this may arise because they allow for mediated as well as non-mediated effects of the intervention but also because they reflect other sources of heterogeneity in the data. Causal analyses can make it possible to pinpoint more exactly what was not well understood and where more data should be brought to bear.

Although many of the specific causal inference methods Zigler and colleagues used were not new, their extensions to air pollution data were a major undertaking in and of themselves. The advances they made in applying the methods to real applications have moved us further than other methodological studies and provided a clearer path toward further development and deployment of the methods in other settings. At the same time, the challenges they faced in both case studies demonstrated the critical importance of involving multidisciplinary teams with detailed technical knowledge of the interventions to make sure that the causal analysis is properly structured, the correct covariates are taken into account, and the results can be correctly interpreted.

The authors argued in their report that “direct-accountability assessments are best equipped to meet the demand of a shifting legislative, judicial, and political environment fraught with questions surrounding the effectiveness of specific policies.” In principle, the Committee agrees that the ability to provide a more complete characterization of the confidence with which one can describe the relationship between an intervention and the outcomes of interest is important both to scientists and to policymakers. What the considerable methodological work in this study indicates, however, is that the presence of a clear causal framework is

not a substitute for detailed consideration of potentially important covariates and the testing of the sensitivity of results to key assumptions made in implementing the methods. Nor can we be sure to what extent the uncertainty in the causal effects estimated is attributable to weakness in the causal relationship or to the imprecision in the definition of the problem and the underlying data. Finally, not all questions can necessarily be addressed in a causal framework, for example, situations in which suitable “controls” do not exist or in which analysts need to predict the potential impacts of some future intervention. The Committee concluded that these and other “direct” accountability methods are an important addition to the “toolkit” and should continue to be further explored, but cannot wholly substitute for “indirect” accountability methods.

---

#### 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 Katy Walker for her assistance in preparing its Critique, to Mary Brennan for science editing of this Report and its Critique, and to Sarah Benckart, Hope Green, Fred Howe, and Ruth Shaw for their roles in preparing this Research Report for publication.

---

#### REFERENCES

---

- Auffhammer M, Bento AM, Lowe SE. 2009. Measuring the effects of the Clean Air Act Amendments on ambient PM<sub>10</sub> concentrations: The critical importance of a spatially disaggregated analysis. *J Environ Econ Manag* 58:15–26.
- Chay K, Dobkin C, Greenstone M. 2003. The Clean Air Act of 1970 and Adult Mortality. *J Risk Uncertainty* 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. *Quart J Econom* 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.
- Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, et al. 2013. Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland. Research Report 176. Boston, MA:Health Effects Institute.
- Harrington W, Morgenstern R, Shih J-S, Bell M. 2012. Did the clean air act amendments of 1990 really improve air quality? *Air Qual Atmos Health*:1–15.
- Harrington W, Morgenstern RD, Shih J-S, Bell M. 2010. Using statistical methods to link stationary source emissions to ambient PM<sub>2.5</sub> concentrations in the eastern United States. Presentation at the 2010 HEI Annual Conference. Health Effects Institute, Boston MA.
- Health Effects Institute. 2010. Proceedings of an HEI Workshop on Further Research to Assess the Health Impacts of Actions Taken to Improve Air Quality. Communication 15. Boston, MA: Health Effects Institute.
- 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.
- HEI Accountability Working Group. 2003. Assessing the Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research. Communication 11. Boston, MA:Health Effects Institute.
- Kelly FJ, Anderson HR, Armstrong B, Atkinson R, Barratt B, Beevers S, et al. 2011. The Impact of the Congestion Charging Scheme on Air Quality in London. HEI Research Report 155. Boston, MA:Health Effects Institute.
- Moore K, Neugebauer R, Lurmann F, Hall J, Brajer V, Alcorn S, et al. 2010. Ambient ozone concentrations and cardiac mortality in Southern California 1983–2000: application of a new marginal structural model approach. *Am J Epidemiol* 171:1233–1243.
- Office of Management and Budget (U.S.). 2010. 2010 Report to Congress on the Benefits and Costs of Federal Regulations and Unfunded Mandates on State, Local, and Tribal Entities. Washington, DC:Office of Information and Regulatory Affairs. Available at [http://www.whitehouse.gov/sites/default/files/omb/legislative/reports/2010\\_Benefit\\_Cost\\_Report.pdf](http://www.whitehouse.gov/sites/default/files/omb/legislative/reports/2010_Benefit_Cost_Report.pdf). Accessed 11/24/2010.
- 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. Boston, MA:Health Effects Institute.
- 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.

- Robins JM, Hernan MA, Brumback B. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550–560.
- Rosenbaum PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55.
- Rubin DB. 2008. For objective causal inference, design trumps analysis. *Ann Appl Stat* 2:808–840.
- Stuart EA. 2010. Matching methods for causal inference: a review and a look forward. *Stat Sci* 25:1–21.
- U.S. Environmental Protection Agency. 1999. The Benefits and Costs of the Clean Air Act 1990 to 2010. EPA Report to Congress. EPA/410/R-99/-001. Washington, DC:Office of Air and Radiation.
- VanderWeele TJ. 2008. Simple relations between principal stratification and direct and indirect effects. *Stat Probabil Lett* 78:2957–2962.
- 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. Boston, MA:Health Effects Institute.
- Zhang J, Zhu T, Kipen H, Wang G, Huang W, Rich D, et al. 2013. Cardiorespiratory Biomarker Responses in Healthy Young Adults to Drastic Air Quality Changes Surrounding the 2008 Beijing Olympics. Research Report 174. Boston, MA:Health Effects Institute.
- Zigler CM, Dominici F, Wang Y. 2012. Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics* 13:289–302.



## RELATED HEI PUBLICATIONS: HEALTH OUTCOMES STUDIES AND STUDIES ASSESSING LONG-TERM HEALTH EFFECTS FROM EXPOSURE TO AIR POLLUTION

Number	Title	Principal Investigator	Date
<b>Research Reports</b>			
178	National Particle Component Toxicity (NPACT) Initiative Report on Cardiovascular Effects	S. Vedal	2013
177	National Particle Component Toxicity (NPACT) Initiative: Integrated Epidemiologic and Toxicologic Studies of the Health Effects of Particulate Matter Components	M. Lippmann	2013
176	Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland	D.W. Dockery	2013
174	Cardiorespiratory Biomarker Responses in Healthy Young Adults to Drastic Air Quality Changes Surrounding the 2008 Beijing Olympics	J. Zhang	2013
170	Impact of the 1990 Hong Kong Legislation for Restriction on Sulfur Content in Fuel	C.-M. Wong	2012
168	Accountability Analysis of Title IV Phase 2 of the 1990 Clean Air Act Amendments	R.D. Morgenstern	2012
163	The London Low Emission Zone Baseline Study	F. Kelly	2011
162	Assessing the Impact of a Wood Stove Replacement Program on Air Quality and Children's Health	C.W. Noonan	2011
155	The Impact of the Congestion Charging Scheme on Air Quality in London	F. Kelly	2011
148	Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes	J.L. Peel	2010
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
137	The Influence of Improved Air Quality on Mortality Risks in Erfurt, Germany	A. Peters	2009
<b>HEI Communications</b>			
15	Proceedings of an HEI Workshop on Further Research to Assess the Health Impacts of Actions Taken to Improve Air Quality		2010
14	HEI's Research Program on the Impact of Actions to Improve Air Quality: 2009 Interim Evaluation and Future Directions		2009
11	Assessing the Health Impact of Air Quality Regulations: Concepts and 2003 Methods for Accountability Research		2003

*Copies of these reports can be obtained from HEI; pdf's are available for free downloading 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 Emeritus and Professor, Civil and Environmental Engineering and Engineering and Public Policy, Carnegie Mellon University*

**Stephen Corman** *President, Corman Enterprises*

**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*

## Health Research Committee

**David L. Eaton, Chair** *Dean and Vice Provost of the Graduate School, University of Washington–Seattle*

**Jeffrey R. Brook** *Senior Research Scientist, Air Quality Research Division, Environment Canada, and Assistant Professor, University of Toronto, Canada*

**David Christiani** *Elkan Blout Professor of Environmental Genetics, Harvard T.H. Chan School of Public Health*

**Francesca Dominici** *Professor of Biostatistics and Senior Associate Dean for Research, Harvard T.H. Chan School of Public Health*

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

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

**Amy H. Herring** *Carol Remmer Angle Distinguished Professor of Children's Environmental Health, and Associate Chair, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina–Chapel Hill*

**Barbara Hoffmann** *Professor of Environmental Epidemiology, Institute for Occupational and Social Medicine, University of Düsseldorf, Germany*

**Allen L. Robinson** *Raymond J. Lane Distinguished Professor and Head, Department of Mechanical Engineering, and Professor, Department of Engineering and Public Policy, Carnegie Mellon University*

**Ivan Rusyn** *Professor, Department of Veterinary Integrative Biosciences, Texas A&M University*

# HEI BOARD, COMMITTEES, and STAFF

## Health Review Committee

**James A. Merchant, Chair** *Professor and Founding Dean Emeritus, College of Public Health, University of Iowa*

**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*

**Jana B. Milford** *Professor, Department of Mechanical Engineering and Environmental Engineering Program, University of Colorado–Boulder*

**Roger D. Peng** *Associate Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

**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*

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

**April Rieger** *Corporate Secretary*

**Zachary Abbott** *Research Assistant*

**Sarah Benckart** *Science Administration Assistant*

**Hanna Boogaard** *Staff Scientist*

**Aaron J. Cohen** *Consulting Scientist*

**Maria G. Costantini** *Principal Scientist*

**Philip J. DeMarco** *Compliance Manager*

**Hope Green** *Publications Associate*

**Anny Luu** *Executive Assistant*

**Hilary Selby Polk** *Managing Editor*

**Tyler Trainor** *Staff Accountant*

**Robert A. Shavers** *Operations Manager*

**Annemoon M.M. van Erp** *Managing Scientist*

**Donna J. Vorhees** *Senior Scientist*

**Katherine Walker** *Senior Scientist*





HEALTH  
EFFECTS  
INSTITUTE

75 Federal Street, Suite 1400  
Boston, MA 02110, USA  
+1-617-488-2300  
[www.healtheffects.org](http://www.healtheffects.org)

RESEARCH  
REPORT

Number 187  
May 2016