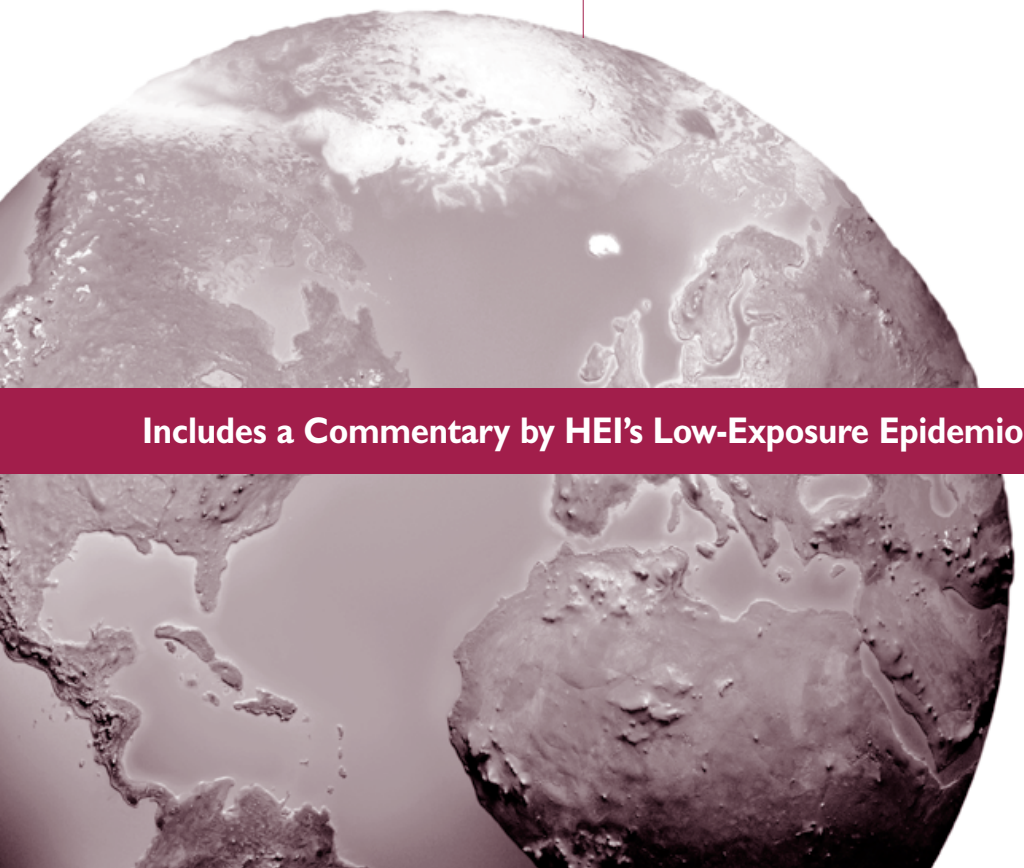
RESEARCH REPORT

HEALTH EFFECTS INSTITUTE

Number 200
November 2019

Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1

Francesca Dominici, Joel Schwartz, Qian Di, Danielle Braun, Christine Choirat, and Antonella Zanobetti



Includes a Commentary by HEI's Low-Exposure Epidemiology Studies Review Panel

ISSN 1041-5505 (print)
ISSN 2688-6855 (online)

Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1

Francesca Dominici, Joel Schwartz, Qian Di, Danielle Braun,
Christine Choirat, and Antonella Zanobetti

with a Commentary by
HEI's Low-Exposure Epidemiology Studies Review Panel

Research Report 200
Health Effects Institute
Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

Publishing history: This document was posted at *www.healtheffects.org* in November 2019.

Citation for document:

Dominici F, Schwartz J, Di Q, Braun D, Choirat C. Zanobetti A. 2019. Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1. Research Report 200. Boston, MA:Health Effects Institute.

 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 of permanent archival quality.

CONTENTS

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 340 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 1,000 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 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. For this study, a special panel — HEI's Low-Exposure Epidemiology Studies Oversight Panel — has worked with the Research Committee in project selection and oversight. The 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. For this study, a special review panel — HEI's Low-Exposure Epidemiology Studies Review Panel — is fulfilling this role.

All project results and accompanying comments by the Review Committee (or, in this case, the Low-Exposure Epidemiology Studies Review Panel) are widely disseminated through HEI's website (*www.healtheffects.org*), printed reports, newsletters and other publications, annual conferences, and presentations to legislative bodies and public agencies.

ABOUT THIS REPORT

Research Report 200, *Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1*, presents a research project funded by the Health Effects Institute and conducted by Dr. Francesca Dominici, of Harvard T.H. Chan School of Public Health, Boston, Massachusetts, and her 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 Low-Exposure Epidemiology Studies Review Panel's comments on the study.

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

The Commentary, prepared by members of the Low-Exposure Epidemiology Studies Review Panel 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 was first examined by outside technical reviewers and a biostatistician. The report and the reviewers' comments were then evaluated by members of the Low-Exposure Epidemiology Studies Review Panel, an independent panel of distinguished scientists who have no involvement in selecting or overseeing HEI studies. During the review process, the investigators had an opportunity to exchange comments with the Review Panel and, as necessary, to revise their report. The Commentary reflects the information provided in the final version of the report.

CONTRIBUTORS

LOW-EXPOSURE EPIDEMIOLOGY STUDIES OVERSIGHT PANEL

Jonathan M. Samet (Chair) *Dean of the Colorado School of Public Health, University of Colorado, Denver*

Amy H. Herring *Sara & Charles Ayres Professor of Statistical Science and Global Health, Duke University, and member of the HEI Research Committee*

Jay H. Lubin *Senior Research Scientist (retired), National Cancer Institute, Division of Cancer Epidemiology & Genetics, Biostatistics Branch*

Fred W. Lurmann *(beginning in 2017) Chairman / Manager of Exposure Assessment Studies, Sonoma Technology, Inc.*

LOW-EXPOSURE EPIDEMIOLOGY STUDIES REVIEW PANEL

Sverre Vedal (Chair) *Professor, Environmental and Occupational Health Sciences, University of Washington, Seattle*

Sara D. Adar *Associate Professor, Epidemiology, School of Public Health, University of Michigan, Ann Arbor*

Benjamin Barratt *Senior Lecturer in Chinese Environment, King's College, London, United Kingdom*

Kiros T. Berhane *Professor of Biostatistics and Director of Graduate Programs in Biostatistics and Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, and member of the HEI Review Committee*

Christopher J. Paciorek *Adjunct Professor, Statistical Computing Consultant, Department of Statistics, University of California, Berkeley*

Jennifer L. Peel *Professor of Epidemiology, Colorado School of Public Health and Department of Environmental and Radiological Health Sciences, Colorado State University, and member of the HEI Review Committee*

Gavin Shaddick *Chair of Data Science and Statistics, Department of Mathematics, University of Exeter, United Kingdom*

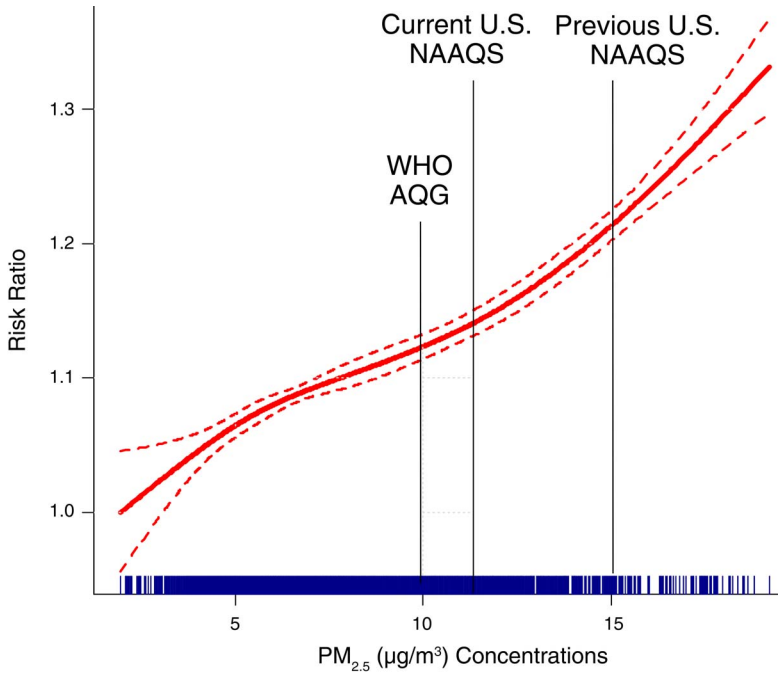
PREFACE

HEI's Program to Assess Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution

INTRODUCTION

Levels of ambient air pollution have declined significantly over the last decades in North America, Europe, and in other developed regions. Despite the decreasing levels of air pollution, recent epidemiological studies report associations between adverse health effects and exposure to air pollution. These studies have found associations between exposure to fine particulate matter, that is, particulate matter ≤2.5 μm in aerodynamic diameter ($PM_{2.5}$*), and mortality at levels below current ambient air quality standards (e.g., Beelen et al. 2014; Crouse et al. 2012; Hales et al. 2012) (Preface Figure 1). In order to improve the science and inform future regulation, it is important to confirm whether associations with adverse health effects continue to be observed as levels of air pollution have declined. It is also important to better understand the shape of the exposure–response function at those low levels.

The growing scientific evidence for effects at levels below current air quality standards and the large overall estimates of the air pollution-attributable burden of



Preface Figure 1. Shape of the concentration–response function for mortality associated with fine particulate matter in a Canadian Cohort. (Courtesy R. Burnett). NAAQS = National Ambient Air Quality Standard; WHO AQG = World Health Organization Air Quality Guidelines.

* A list of abbreviations and other terms appears at the end of this volume.

disease, as well as the interest in reducing greenhouse gases, suggest that more stringent air quality standards and guidelines may be considered in the future. For these reasons, there is a need for additional investigation to improve our understanding of exposure–response function(s) for mortality and morbidity at low levels of $PM_{2.5}$, ozone (O_3), and other ambient air pollutants. Such studies would inform risk assessors and policy makers regarding exposure–response functions at levels of ambient air pollution currently prevalent in North America, Western Europe, and other high-income regions of the world.

In 2014, HEI issued RFA 14-3, *Assessing Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution,* to solicit studies to address these important questions. The main goals of the RFA were to:

1. Fund studies to assess health effects of long-term exposure to low levels of ambient air pollution, including all-cause and cause-specific mortality and morbidity. Such studies should analyze and evaluate exposure–response function(s) for $PM_{2.5}$ and other pollutants at levels currently prevalent in North America, Western Europe, and other high-income regions. The studies may also address related questions about health effects at low levels of ambient air pollution.
2. Develop statistical and other methodology required for, and specifically suited to, conducting such research including, but not limited to, evaluation and correction of exposure measurement error.

Applicants were asked to pay particular attention to having sufficiently large cohorts and statistical power to detect associations should they exist, having the ability to test various potential confounders of any associations, and developing exposure-assessment approaches and statistical methodology that would enable a robust examination of the associations.

Specifically, applicants were asked to propose studies to:

1. Compare and contrast alternative analytic models and accompanying uncertainty. For example, compare threshold against nonthreshold models, linear against nonlinear models, and parametric against nonparametric models, to characterize the exposure–response function(s) at low levels of ambient air pollution.
2. Explore possible variability in estimates of risk at low pollutant concentrations among populations, and identify possible contributing factors. Such factors could include age, smoking, socioeconomic position, health status, and access to medical care, as well as differences in air pollution sources and time–activity patterns.
3. Develop and evaluate exposure-assessment methods suitable to estimate exposure to low levels of air pollution at various spatial and temporal scales in large study populations, including people who reside in areas not covered by routine ground-level monitoring.
4. Develop, evaluate, and apply statistical methods to quantify and correct for exposure measurement error in risk estimates and in characterization of exposure–response relationships.
5. Develop and validate approaches to assess the effects of co-occurring pollutants on any health effect associations at low ambient concentrations.
6. Develop and validate indirect approaches to correct risk estimates for the effects of important potential confounding variables, such as smoking, in the absence of such data at the individual level.
7. Improve techniques for record linkage and methods for disclosure protection for optimal use of large administrative databases in air pollution and health research.

STUDY SELECTION

HEI established an independent Low-Exposure Epidemiology Studies Oversight Panel — consisting of outside experts and HEI Research Committee members — to prepare RFA 14-3 and review all applications submitted in response (see Contributors page). Members of HEI's Research Committees with any conflict of interest were recused from all discussions and from the decision-making process. The HEI Research Committee reviewed the Panel's recommendations and recommended three studies for funding to HEI's Board of Directors, which approved funding in 2015.

Preface

This Preface summarizes the three studies, HEI's oversight process, and the review process for the Phase 1 reports.

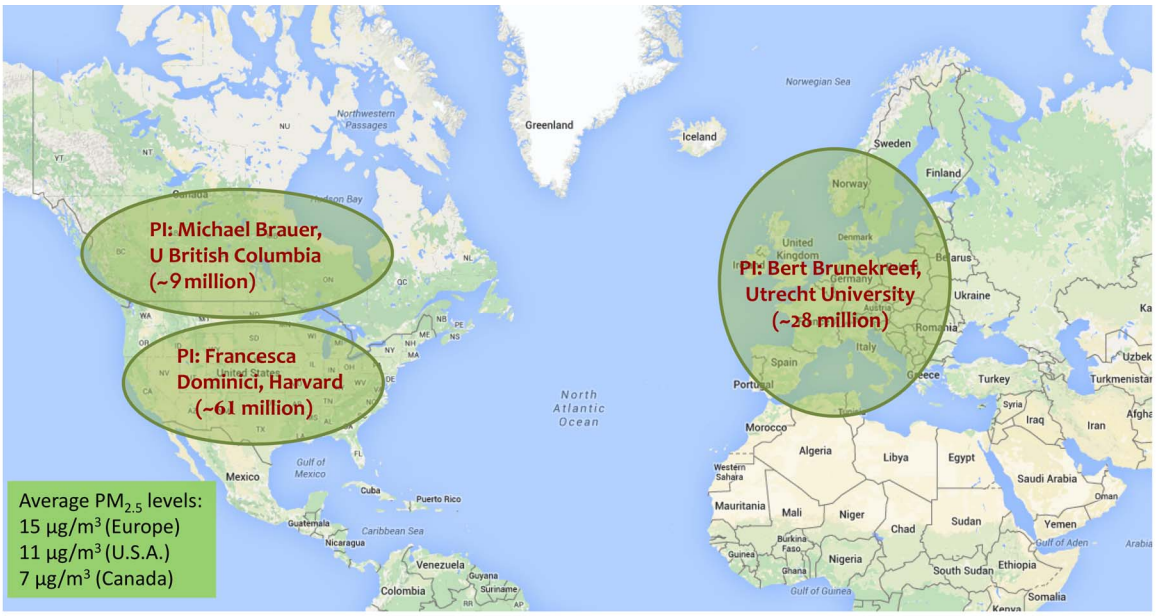
OVERVIEW OF THE HEI LOW-EXPOSURE EPIDEMIOLOGY STUDIES

After a rigorous selection process, HEI funded three teams, led by Michael Brauer at The University of British Columbia, Canada, Francesca Dominici at the Harvard T.H. Chan School of Public Health, United States, and Bert Brunekreef at the University of Utrecht, the Netherlands, to investigate health effects of exposure to low levels of air pollution in very large populations in Canada, the United States, and Europe, respectively (see Preface Table and Preface Figure 2). The studies included large population cohorts (with detailed individual information about potential confounders for all subjects or for subsets of cohorts), as well as large administrative databases with greater statistical power (albeit with less individual information about potential confounders). Additionally, the three teams employed satellite data and ground-level pollutant measurements, used high-quality exposure-assessment models at high spatial resolutions, and set out to develop and apply novel statistical methods.

The three studies are expected to inform the scientific community and the risk assessors and policy makers regarding exposure–response functions at levels of ambient air pollution currently prevalent in North

Preface Table. HEI's Program to Assess Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution

Investigator (Institution)	Study Title	Phase 1 Report	Phase 2 (Final) Report Expected
Brauer, Michael (The University of British Columbia, Canada)	Mortality–Air Pollution Associations in Low Exposure Environments (MAPLE)	Brauer M, Brook JR, Christidis T, Chu Y, Crouse DL, Erickson A, et al. 2019. Mortality–Air Pollution Associations in Low-Exposure Environments (MAPLE): Phase 1. Research Report 203. Boston, MA:Health Effects Institute.	Summer 2020
Brunekreef, Bert (Utrecht University, the Netherlands)	Mortality and Morbidity Effects of Long-Term Exposure to Low-Level $PM_{2.5}$, Black Carbon, NO_2 and O_3: An Analysis of European Cohorts	None	Fall 2020
Dominici, Francesca (Harvard University, T.H. Chan School of Public Health, U.S.A.)	Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Pollution	Dominici F, Schwartz J, Di Q, Braun D, Choirat C, Zanobetti A. 2019. Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1. Research Report 200. Boston, MA:Health Effects Institute.	Summer 2020



Preface Figure 2. Geographical areas and populations covered by HEI's research program to assess adverse health effects of long-term exposure to low levels of ambient air pollution.

America, Western Europe, and other developed regions. The full sets of analyses are expected to be completed in 2020, as discussed in the following sections.

CANADIAN STUDY (MICHAEL BRAUER ET AL.)

Brauer and colleagues are assessing the relationship between nonaccidental mortality and long-term exposure to low concentrations of $PM_{2.5}$ in four large population-based cohorts, including a careful characterization of the shape of the exposure–response function. The investigators are using Canadian census data and have access to a nationally representative population of approximately 9 million Canadians (ages 25–90 yr) (Preface Figure 2). The Canadian team is developing hybrid models primarily using satellite data, as well as chemical transport models, land-use variables, and routinely collected monitoring data for $PM_{2.5}$. They are also estimating ambient concentrations for nitrogen dioxide (NO_2) and O_3 for Canada and the United States during the period 1981–2016. Additionally, they will be validating satellite data against ground-based monitors in Canada as part of the SPARTAN network (Snider et al. 2015).

The exposure models are applied to estimate effects of air pollution exposure on all-cause and cause-specific mortality in four Canadian cohorts:

1. About 2.5 million respondents who completed the 1991 census long form of the Canadian Census Health & Environment Cohorts (CanCHEC),
2. About 3 million respondents who completed the 1996 CanCHEC census long-form,
3. About 3 million respondents who completed the 2001 CanCHEC census long-form, and
4. About 540,000 respondents who participated in the Canadian Community Health Survey (CCHS) between 2001 and 2012, and reported individual-level risk factors, including smoking.

EUROPEAN STUDY (BERT BRUNEKREEF ET AL.)

Brunekreef and colleagues are basing their study on the European Study of Cohorts for Air Pollution Effects (ESCAPE), which started about a decade ago; its results have been published widely (e.g., Beelen et al. 2014). In the current HEI-funded study, the investigators are analyzing pooled data from 10 ESCAPE cohorts (instead of

the cohort-specific approach they used previously). In addition, they are using data from six large administrative cohorts to yield a total study population of approximately 28 million Europeans (Preface Figure 2). They are developing hybrid, Europe-wide and location-specific exposure models that utilize land-use information, dispersion modeling, satellite data, ESCAPE monitoring data, and routinely collected monitoring data for $PM_{2.5}$, NO_2, O_3, and black carbon at high spatial resolution (residential address level; such detailed information is very difficult to obtain in the United States).

Brunekreef and colleagues are investigating the following health outcomes: all-cause and cause-specific mortality, incidence of coronary and cerebrovascular events, and lung cancer incidence. The incorporation of ESCAPE cohorts with individual covariate information as well as very large administrative cohorts (albeit with less detailed information) will provide new insights in the merits of both approaches.

UNITED STATES STUDY (FRANCESCA DOMINICI ET AL.)

Dominici and colleagues are evaluating Medicare and Medicaid data for a study population of approximately 61 million Americans (Preface Figure 2). They are developing high spatial resolution (1 km^2-grid) hybrid exposure models that incorporate satellite data, chemical transport models, land-use and weather variables, and routinely collected monitoring data for NO_2, O_3, and $PM_{2.5}$ and its components, for the continental United States during the period 2000–2012. Exposure models will be applied to estimate adverse health effects of air pollution in three cohorts:

1. Medicare enrollees (28.6 million elderly enrollees per year, 2000–2012);
2. Medicaid enrollees (28 million enrollees per year, 2010–2012); and
3. Medicare Current Beneficiary Survey enrollees (nationally representative sample of approximately 15,000 enrollees per year with rich individual-level risk factor information, including smoking).

Dominici and colleagues are analyzing the following health outcomes: time to death, time to hospitalization by cause, and disease progression (time to rehospitalization). They are developing and applying new causal inference methods to estimate exposure–response functions to adjust for confounding and exposure measurement error. Additionally, they are developing tools for reproducible research including approaches for data sharing, record linkage, and statistical software.

STUDY OVERSIGHT

HEI's independent Low-Exposure Epidemiology Studies Oversight Panel continues to provide advice and feedback on the study design, analytical plans, and study progress throughout the duration of the research program.

Given the substantial challenges in conducting a systematic analysis to assess health effects of long-term exposure to low levels of ambient air pollution, HEI has worked actively (and continues to do so) with the study teams to coordinate their efforts and ensure the maximum degree of comparable epidemiological results at the end of this research effort. To this end, HEI has regularly held investigator workshops and site visits, among other activities. In addition, the studies are subject to HEI's special Quality Assurance procedures, which include an audit by an independent audit team (see *www.healtheffects.org/research/quality-assurance).*

REVIEW OF PHASE 1 AND PHASE 2 (FINAL) REPORTS

To inform the ongoing review of the U.S. National Ambient Air Quality Standards (NAAQS) for $PM_{2.5}$ and O_3 starting in 2018, HEI requested Phase 1 reports from the investigators based on the research completed during the first two years of the Canadian and U.S. studies. Thus, the Phase 1 reports by Drs. Brauer and Dominici provide summaries of results to date, including those published in journal articles.

As is common for major research programs, HEI convened a Low-Exposure Epidemiology Studies Review Panel to independently review the Phase 1 reports by Drs. Brauer and Dominici. The Panel consists of seven experts in epidemiology, exposure assessment, and biostatistics (see Contributors page). Commentaries by the Review Panel accompany the Phase 1 reports. The Panel will also review the final reports of the three studies.

The Phase 1 Research Reports provide an opportunity to present the results from the first two years of research in one place and to present the Review Panel's Commentaries, which review the results and evaluate the studies' strengths and weaknesses. The three studies commenced in spring 2016 and are expected to be completed in summer 2020, with final reports published during 2021.

REFERENCES

Beelen R, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, Hoffmann B, et al. 2014. Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. Lancet 383:785–795.

Crouse DL, Peters PA, van Donkelaar A, Goldberg MS, Villeneuve PJ, Brion O, et al. 2012. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. Environ Health Perspectives 120:708–714.

Hales S, Blakely T, Woodward A. 2012. Air pollution and mortality in New Zealand: Cohort study. J Epidemiol Community Health 66:468–473.

Snider G, Weagle CL, Martin RV, van Donkelaar A, Conrad K, Cunningham D et al. 2015. SPARTAN: A global network to evaluate and enhance satellite-based estimates of ground-level particulate matter for global health applications. Atmos Meas Tech Discuss 7:7569–7611.

HEI STATEMENT

Synopsis of Research Report 200

Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1

INTRODUCTION

The levels of most ambient air pollutants have declined significantly in the United States during the last few decades. Recent epidemiological studies, however, have suggested an association between exposure to ambient levels of air pollution — even below the current U.S. National Ambient Air Quality Standards (NAAQS) — and adverse

What This Study Adds

- This study is part of an HEI program to address questions regarding potential associations between air pollution exposure and health outcomes at low ambient air pollution levels, particularly at levels below the current U.S. national air quality standards.
- Dominici and colleagues developed hybrid, U.S.-wide models using machine learning to estimate outdoor fine particle (particulate matter ≤ 2.5 μm in aerodynamic diameter, or $PM_{2.5}$) and ozone (O_3) concentrations at 1 km $\times$ 1 km grids, by combining monitoring, satellite, transport modeling output, and other data.
- They obtained Medicare data for 61 million Americans, ages 65 years and older, who enrolled between 2000 and 2012. Using both cohort and case–crossover designs, they analyzed the association between long-term and short-term outdoor $PM_{2.5}$ and O_3 exposures and mortality.
- The investigators report positive associations between nonaccidental, all-cause mortality and $PM_{2.5}$ and O_3 at low concentrations, including below the U.S. National Ambient Air Quality Standards (annual 12 $\mu g/m^3$ for $PM_{2.5}$ and 8-hour 70 ppb for O_3).
- These associations were robust to most adjustments for potential confounding by a number of lifestyle and behavioral factors in the cohort analyses. Sensitivity analyses did not meaningfully impact the findings of association.
- HEI's Low-Exposure Epidemiology Studies Review Panel noted, however, that several important issues still need to be addressed by the investigators regarding these results during the remainder of this project. In particular, the potential for confounding by time and the complexities introduced by the use of different spatial scales for the exposure and health data need to be explored in more detail, and the causal inference methods need to be more fully applied.
- The Panel concluded that Dominici and colleagues have conducted an extensive and innovative set of initial analyses in these extraordinarily large air pollution and health data sets. While initial conclusions may be drawn from these analyses, the Panel awaits the further analyses that are underway before reaching full conclusions on the air pollution and public health implications of this important research.

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. Francesca Dominici at the Harvard T.H. Chan School of Public Health, Boston, Massachusetts, and colleagues. Research Report 200 contains both the detailed Investigators' Report and a Commentary on the study prepared by HEI's Low-Exposure Epidemiology Studies Review Panel.

health effects. In view of the importance of such research findings, the Health Effects Institute in 2014 issued a request for applications (RFA 14-3) seeking to fund research to assess the health effects of long-term exposure to low levels, particularly below the NAAQS, of ambient air pollution and to develop improved statistical methods for conducting such research. HEI funded three studies under this program; each study used state-of-the-art exposure methods and very large cohorts. The studies were based in the United States, Canada, and Europe, thus providing a comprehensive cross-section of high-income countries where ambient levels are generally low.

The low-exposure-level studies are scheduled to be completed in 2020. In 2018, in order to inform the ongoing review of the NAAQS for fine particles ($PM_{2.5}$) and ozone (O_3), HEI requested Phase 1 reports from the U.S. (Francesca Dominici) and Canadian (Michael Brauer) investigators. HEI's formed a special panel, the Low-Exposure Epidemiology Studies Review Panel, to evaluate the studies' methods, results, conclusions, and their strengths and weaknesses. This Statement focuses on the study by Dr. Francesca Dominici, from the Harvard T.H. Chan School of Public Health, Boston, Massachusetts, and her colleagues, titled, "Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution."

APPROACH

Aims: The aims of the Dominici study were to (1) develop hybrid, high-resolution, exposure-prediction models to estimate long-term exposures to $PM_{2.5}$ and O_3 levels for the continental United States; (2) develop and apply causal inference methods; (3) estimate all-cause mortality associated with exposure to ambient air pollution for all U.S. Medicare enrollees between 2000 and 2012 using a cohort (long-term) and a case–crossover (short-term) design; and (4) develop tools for data sharing, record linkage, and statistical software.

Data and Methods: Dominici and colleagues developed hybrid air pollution concentration models for the contiguous United States for the period 2000 to 2012, using data from a variety of sources, including satellite data, chemical transport models, land-use and weather variables, and routinely collected air monitoring data from the U.S. Environmental Protection Agency (EPA).

With this large amount of data and using multiple approaches and input variables, the investigators developed a hybrid model to estimate daily $PM_{2.5}$ and O_3 concentrations at 1 km $\times$ 1 km grids across the continental United States. Complex atmospheric processes were addressed using a neural network that modeled nonlinearity and interactions. The neural network was trained using data covering the study period, and the predictions were validated against 10% of the EPA air monitors left out of the model. A similar approach was used to estimate and validate a model to predict O_3 concentrations during the warm months (April through September) of each study year.

Health data were obtained from the Centers for Medicare and Medicaid Services for all Medicare enrollees for the years 2000 to 2012, which represents more than 96% of the U.S. population 65 years of age and older (see Statement Table). The study obtained records for all Medicare enrollees (~61 million), with 460 million person-years of follow-up and 23 million deaths. They also obtained covariate information from the Medicare Current Beneficiary Survey (MCBS; ~57,000 people), an annual phone survey of a nationally representative sample of Medicare beneficiaries, with information on more than 150 individual-level risk factors, including smoking and body mass index.

Using the Medicare data and cohort and case–crossover designs, they investigated the association between exposure to $PM_{2.5}$ and O_3 and all-cause mortality in two-pollutant analyses, including separate analyses for low pollutant concentrations. For the cohort study, they performed survival analyses using the Andersen-Gill method, a variant of the traditional Cox proportional hazards model that incorporates spatiotemporal features by allowing for variation in covariates from year to year. The investigators developed concentration–response curves by fitting a log-linear model with thin-plate splines for both pollutants while controlling for important individual and ecological variables, including socioeconomic status and race. For the case–crossover study, the case day was defined as the date of death, with exposure defined as the mean of the ambient concentration on that day and the day before; this was compared to exposure on three predefined control days. They fitted a conditional logistic regression to all pairs of case and matched control days, thus estimating the relative risk of all-cause mortality associated

Statement Table. Key Features of the Dominici et al. Study

Overall	
Medicare study population	60.9 million
MCBS study population	57,200
Study period	2000–2012
Case–Control Study	
Follow-up period	460.3 million person-years
Deaths	22.6 million
$PM_{2.5}$ average concentration	11.0 μg/m^3
O_3 average concentration	46.3 ppb
Case–Crossover Study	
Case days	22.4 million
Control days	76.1 million
$PM_{2.5}$ average concentration	11.6 μg/m^3
O_3 average concentration	37.8 ppb

with short-term exposure to $PM_{2.5}$ and O_3. They also performed subanalyses to explore the health effects at lower levels of exposure.

To assess whether any subgroups within the cohort study were at higher or lower risk of mortality associated with either long-term or short-term air pollution exposure, the investigators fitted the same statistical models to certain population subgroups (e.g., male vs. female and white vs. black). To explore the robustness of the results from the cohort analysis, they performed sensitivity analyses and compared any changes in risk estimates with differences in confounder adjustment and estimation approaches. Finally, since Medicare data do not include information on many important individual-level covariates, the investigators utilized data from the Medicare Current Beneficiary Statement to examine how the lack of adjustment for these risk factors could have affected the risk estimates for the Medicare cohort.

RESULTS

Dominici and colleagues report overall good performance of the models for estimating $PM_{2.5}$ and O_3 concentrations, with overall R^2 values of 0.84 and 0.80, respectively. For $PM_{2.5}$, the average annual concentration was 11.0 μg/m^3 during the study period, 2000–2012. Performance of the model varied between different geographical regions and seasons; the highest $PM_{2.5}$ concentrations were predicted to be in California and the eastern and southeastern United States, and model performance was better in the eastern and central United States than in the western part of the country. And, the $PM_{2.5}$ model performed best during the summer. For O_3, the average of 8-hour daily concentrations during the warm season was 46.3 ppb during the study period. O_3 concentrations were highest in the Mountain region and in California and lower in the eastern states. The average concentrations of $PM_{2.5}$ decreased during the study period, but O_3 concentrations remained more or less the same. Annual $PM_{2.5}$ and warm-season O_3 concentrations were only weakly correlated.

The 2000–2012 cohort of Medicare beneficiaries provided a very large population for studying association with long-term effects of exposure to ambient air pollution. In two-pollutant analyses of long-term effects, Dominici and colleagues report a 7.3% higher risk of all-cause mortality for each 10-μg/m^3 increase in annual average $PM_{2.5}$ concentrations and a 1.1% higher risk of mortality for each 10-ppb increase in average O_3 concentrations in the warm season. At low concentrations — less than 12 μg/m^3 $PM_{2.5}$ and less than 50 ppb O_3 — the risk was 13.6% for $PM_{2.5}$ and 1.0% for O_3 for each 10-μg/m^3 and 10-ppb increase in concentrations, respectively. The concentration–response relationships from the two-pollutant models showed almost

linear curves, with no suggestion of a threshold down to 5 μg/m^3 $PM_{2.5}$ and 30 ppb O_3.

In subgroup analyses for long-term $PM_{2.5}$ exposure, the investigators found larger estimates of effect among males and among Hispanics, Asians, and particularly African Americans, compared with whites. Individuals with low socioeconomic status, as indicated by eligibility for Medicaid, appear to have a slightly higher risk per unit of $PM_{2.5}$ exposure. For long-term O_3 exposure, the subgroup analysis showed that the effect estimates were higher for Medicaid-eligible enrollees and slightly higher for whites, but these analyses also produced puzzling hazard ratios of less than 1 for certain subgroups, including Hispanics and Asians, and particularly for Native Americans, than the overall population.

For short-term exposures, the investigators observed a 1.05% greater risk of mortality in two-pollutant models for a 10-μg/m^3 increase in $PM_{2.5}$ concentrations and a 0.51% greater risk for a 10-ppb increase in 8-hour warm-season O_3 concentration. (Pollutant levels were averaged over the current and previous day.) At low concentrations (below 25 μg/m^3 of $PM_{2.5}$ and below 60 ppb of O_3), the associations remained elevated for both pollutants (1.61% for $PM_{2.5}$ and 0.58% for O_3). The concentration–response curves showed the relative risk increasing sharply for both pollutants at a relatively low concentration and then leveling out at higher concentrations. The investigators observed evidence of effect modification for several variables, including a higher $PM_{2.5}$–mortality risk for females than for males.

INTERPRETATION OF RESULTS

In its independent review of the study, HEI's Low-Exposure Epidemiology Studies Review Panel noted that the report by Dominici and colleagues summarizes an impressive amount of work completed in the first part of this HEI project. Particularly strong aspects of this work include the extremely large, national cohort, with high-resolution exposure assessment and development and application of state-of-the-art statistical techniques. The Panel also noted that additional research, including further development of causal methods that would properly allow for the complexities in the design of the studies and nature of the data, is currently ongoing.

Exposure Assessment: The use of large, diverse, and existing data sets to generate estimates of $PM_{2.5}$ and O_3 concentrations on a 1 km × 1 km national grid for the entire continental United States (~8 million km^2) is impressive, and allowed the investigators to estimate concentrations in areas where air monitors are sparse. However, as with any exposure assessment, it is critical to consider the potential for exposure prediction errors.

Despite steps to correct for regional and compositional differences, both geographical and temporal variability in the errors of the concentration estimates persisted in the final estimates for $PM_{2.5}$ and O_3. The exposure model was trained by leaving out 10% of EPA air quality monitors. But because these monitors are generally located in areas with high population density, it is possible that the model is prone to larger error in areas with lower population density — which generally have lower $PM_{2.5}$ concentrations and therefore are of greater interest in the context of this study. And, based on earlier work by the researchers that provides the basis for the exposure models used in these studies, it appears that the model may systematically underpredict concentrations for unexplained reasons. The nature, sources, size, and potential impact of the potential errors discussed here are important to understand and deserve attention in future analyses.

Long-Term Health Effects, Cohort Study: Using the massive database of all Medicare recipients during 2000 to 2012, and combining it with the equally large exposure predictions, Dominici and colleagues have performed a study with extraordinary statistical power to investigate the association between all-cause mortality and long-term exposure to a range of $PM_{2.5}$ and O_3 levels. That they observed an association between annual average concentrations and mortality at higher concentrations was not the new finding of this research, but the findings at low levels, particularly at levels below the current NAAQS, are novel and potentially important.

The greatest challenge to the internal validity of this study, as for all observational studies, is the potential for confounding, which can bias the results. To address such concerns, the investigators performed numerous analyses with some 20 covariates. They also utilized findings from a smaller Medicare cohort that had a much richer set of potential confounding variables to assess the likely impact of having only a limited number of covariates in the main cohort analysis. In addition, to allow for the effects of time-dependent covariates known to vary from year to year, they utilized a variant of the classic Cox proportional hazards model, the Andersen-Gill formulation.

However, this is a complex study. Health and personal characteristics are available for individuals, but ambient air pollutant exposure is estimated at the ZIP code level (averaged from the 1 km × 1 km spatial scale of the prediction model). Additionally, the ZIP code scale is the smallest spatial unit at which individual residential and other covariate information is available. These factors, coupled with confounders that can act at the level of the individual, the community, or the regional environment, result in a complex hybrid model. These issues pose important challenges for the next phase of the work planned by the investigators, and the causal inference methods under development will need to focus on these challenges.

Based on the current results, the Panel offers the following comments most relevant to the cohort analyses.

The investigators performed various analyses to explore the potential impact of confounding; however, the Panel noted several areas with a potential for residual confounding in the cohort study. For example, some results from the subgroup analyses are puzzling, particularly the dramatically higher effect of $PM_{2.5}$ exposure in African Americans and the negative (protective) effects of exposure to O_3 for Native Americans, Hispanics, and Asians.

Although the investigators have used the Andersen-Gill formulation to better model time-dependent variables, the Panel's biggest concern relates to the problem of potential for temporal confounding, with both overall nonaccidental mortality and $PM_{2.5}$ levels declining steadily over the period of the study, 2000–2012. Because this is an open cohort (new individuals enter the cohort as they enroll for Medicare), age — which is controlled in the analyses — is not necessarily strongly correlated with calendar time. As a result, confounding could occur because of the contributions of both age and calendar time. The Panel believes that without accounting for confounding by time, the findings of the long-term exposure study should be viewed with caution.

The Panel also has concerns about the impact of the likely exposure misclassification and confounding related to the hybrid nature of the study, but appreciates that exposure measurement error correction methodology for spatially varying pollutants and methods to address confounding in such a complex study setting are still in their infancy. Additionally, the Panel notes that data on individual health-related behaviors, such as smoking, diet, and exercise, do not capture the full extent of variability in the behaviors, such as geographical variability. Finally, the presence of other pollutants — such as NO_2 — may also confound the associations between $PM_{2.5}$ and O_3 and mortality.

Another important issue in interpretation of these results is related to the very large population studied here, and consequently the very high apparent precision of the results (i.e., the very small confidence intervals). Because the impact of bias and model misspecification is not reflected in standard uncertainty measures, one should be cautious about over-interpreting the narrow confidence intervals. The Panel's comments and concerns about the potential impacts of bias and of unmeasured confounding should be viewed in this broader context.

Short-Term Health Effects, Case-Crossover Study: The second study in this report uses a case–crossover design — a variant of the time-series design — to evaluate short-term effects of low-level air pollution in the Medicare population. One advantage this study design has over the long-term design is that it is based on variation in exposure and mortality experienced by an individual over short periods of time (days, rather than years). Therefore, only confounding factors that vary over short periods of time, such as weather, are of potential concern, rather than the much larger array of potential confounders that either do not vary with time or have long-term trends. On the other hand, by design, time-series analyses only address the immediate impact of air pollution on mortality rather than the pollutants' role in the development of chronic morbidity and subsequent mortality.

Dominici and colleagues report a relative risk increase of 1.05% and 0.51% in daily mortality rate for each 10-$\mu g/m^3$ increase in $PM_{2.5}$ and 10-ppb increase in O_3, respectively. The concentration–response analyses for $PM_{2.5}$ and O_3 suggest a nonlinear relationship, with a steeper slope at low concentrations and flattening at higher concentrations. They have also investigated effect modifications for a range of variables. For example, they report that the mortality effect of short-term exposure to $PM_{2.5}$ is greater in women than in men, in contrast to the finding in the cohort study. The effects in other subgroup analyses were generally not significant, except for Medicaid eligibility. Also, NO_2 — another time-varying covariate — was not included in these analyses.

Causal Modeling: There is increasing interest in research on casual inference methods because of the challenges in accounting for confounding in the preceding analyses of observational data, and Dominici and colleagues are devoting significant effort to the development and extension of two such methods.

In the first method, the investigators have developed a generalized-propensity-score approach for confounding adjustment along with a regression calibration method to address exposure measurement error in health models. In the second approach, they have developed a new Bayesian causal approach, known as *local exposure–response confounding adjustment*, to estimate exposure–response curves accounting for differential effects of confounders at different levels of exposure. Both of these approaches serve as potentially useful starting points, and the Panel notes that current applications do not address the concerns raised about the long-term and short-term studies — in particular, concerns about residual confounding and impacts of the complex hybrid nature of the study designs — and so it looks forward to the full development and applications of these methods to the health analyses.

Sharing of Models and Data: Dominici and colleagues have made a special effort to make available their data, workflows, and analyses, and have posted these at a secure high-performance computing cluster with the objective of developing an open science research data platform. Additionally, the codes and software tools are publicly available from another depository. The investigators' work in these areas will continue. The Panel finds these efforts praiseworthy and encourages the Dominici team to continue sharing the unique resources they have developed.

CONCLUSIONS

Using very large air pollution model and health data sets, Dominici and colleagues have reported initial results using two types of analysis — a cohort analysis of long-term exposures and a case–crossover analysis of short-term exposures. They and found positive associations of both $PM_{2.5}$ and O_3 with all-cause mortality, with associations extending to concentrations below the current NAAQS and with little evidence of a threshold. The investigators also conducted a range of sensitivity analyses and controlled for many confounders; these did not meaningfully change the initial findings of associations. These initial analyses are thorough and comprehensive, and make a valuable contribution to the literature.

As extensive as these analyses are, as noted by the Panel and by the investigators, there are several key questions that need to be investigated further before firmer conclusions can be drawn. Particularly important among these are (1) issues around the potential for confounding by time trends and other variables, including other pollutants such as NO_2, and geographical patterns in exposure and health status; (2) impact of the different spatial scales of the variables in both the long-term and short-term analyses, and the resulting complex quasi-ecological (hybrid) nature of the models, with the potential for exposure misclassification and residual confounding; and (3) extension of their work by the development, testing, and application of causal inference methods in the full study population.

Dominici and colleagues have performed a set of extensive and creative analyses in the largest air pollution and health databases to date. While initial conclusions may be drawn from these first analyses, the Panel will wait for the planned extensive further analyses to be completed before reaching full conclusions on the air pollution and public health implications of this important research.

INVESTIGATORS' REPORT

Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1

Francesca Dominici[1], Joel Schwartz[1], Qian Di[1], Danielle Braun[1,2], Christine Choirat[1], and Antonella Zanobetti[1]

[1]Harvard T.H. Chan School of Public Health, Boston, Massachusetts; [2]Department of Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts

ABSTRACT

Introduction. This report provides a summary of major findings and key conclusions supported by a Health Effects Institute grant aimed at "Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Pollution." Our study was designed to advance four critical areas of inquiry and methods development.

Methods. First, our work focused on predicting short- and long-term exposures to ambient $PM_{2.5}$* mass (particulate matter $\leq 2.5\mu m$ in aerodynamic diameter) and ozone (O_3) at high spatial resolution (1 km × 1 km) for the continental United States during the period 2000–2012 and linking these predictions to health data. Second, we developed new causal inference methods for exposure–response (ER) that account for exposure error and adjust for measured confounders. We applied these methods to data from the New England region. Third, we applied standard regression methods using Medicare claims data to estimate health effects that are associated with short- and long-term exposure to low levels of ambient air pollution. We conducted sensitivity analyses to assess potential confounding bias due to lack of extensive information on behavioral risk factors in the Medicare population using the Medicare Current Beneficiary Survey (MCBS) (nationally representative sample of approximately 15,000 Medicare enrollees per year), which includes abundant data on individual-level risk factors including smoking. Finally, we have begun developing tools for reproducible research — including approaches for data sharing, record linkage, and statistical software.

Results. Our HEI-funded work has supported an extensive portfolio of analysis and the development of statistical methods that can be used to robustly understand the health effects of long- and short-term exposure to low levels of ambient air pollution. This report provides a high-level overview of statistical methods, data analysis, and key findings, as grouped into the following four areas: (1) Exposure assessment and data access; (2) Epidemiological studies of ambient exposures to air pollution at low levels; (3) Methodological contributions in causal inference; and (4) Open science research data platform.

Conclusion. Our body of work, advanced by HEI, lends extensive evidence that short- and long-term exposure to $PM_{2.5}$ and O_3 is harmful to human health, increasing the risks of hospitalization and death, even at levels that are well below the National Ambient Air Quality Standards (NAAQS).

This Investigators' Report is one part of Health Effects Institute Research Report 200, which also includes a Commentary by the Institute's Low-Exposure Epidemiology Studies Review Panel and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. Francesca Dominici, Harvard T.H. Chan School of Public Health, 677 Huntington Ave., Boston, MA 02115; e-mail: *fdominic@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.

* A list of abbreviations and other terms appears at the end of this volume.

INTRODUCTION

In late 2014, HEI issued a Request for Applications (RFA 14-3) seeking proposals to assess health effects of long-term exposure to low levels of ambient air pollution with particular attention to (a) sufficient size and statistical power to detect associations if they exist, (b) the ability to test different potential confounders of these associations, and (c) a variety of approaches to exposure assessment and statistical analysis to enable a robust examination of the associations.

Levels of ambient air pollution have declined significantly over the last decades in North America, Europe, and in other developed regions. Nonetheless, epidemiological studies continue to report associations of adverse health effects with air pollution even at these lower levels, and recently some studies have found associations at levels below current ambient air quality standards (e.g., Crouse et al. 2012; Hales et al. 2012; Shi et al. 2016). In order to inform future risk assessment and regulation, HEI committed funding to examine whether associations with adverse effects continue to be observed as levels of air pollution decline further and what the shape of the ER function is at those low levels, both major uncertainties in current air quality standards decision making.

As air pollution levels continue to decrease and regulatory actions become more costly, the quantification of the public health benefits of cleaner air will be subject to an increased level of scrutiny. Epidemiological analyses of claims data have provided strong evidence of air pollution's adverse health effects, mostly using data from urban areas (Carey et al. 2013; Crouse et al. 2015; Krewski et al. 2009; Ostro et al. 2015; Turner et al. 2016). Yet, significant gaps in knowledge remain, particularly with regard to the health effects of long-term exposure to lower levels of air pollution, and no large study to date has investigated the health effects of long-term air pollution in areas with sparse monitoring (Aim 1).

The estimation of health effects associated with long-term exposure to low levels of air pollution presents key methodological challenges, including: (1) the estimation of an ER within a traditional regression framework does not have a causal interpretation and can be highly sensitive to model choice for both the shape of the ER and the adjustment for confounding; (2) health effects estimation at low exposure levels might be affected by a different set of confounders than at high exposure levels; (3) information on individual-level potential confounders is limited in the administrative data; (4) estimation of the ER must account for potentially larger exposure error at lower exposure levels; (5) identification of effect modifiers is challenged by the large number of possibilities that cannot all be tested individually; and (6) causal estimation of ER in the context of multiple pollutants is virtually nonexistent in the literature. A rigorous treatment of *all* these statistical challenges, under a unifying causal inference framework, is necessary to investigate the health risks associated with low pollution levels and to inform regulatory policy (Aim 2). Additionally, little is known about health effects at low pollution levels, not only on mortality and morbidity outcomes, but also on disease progression in populations thought to be highly susceptible to air pollution — such as low-income adults (Aim 3). Finally, methods for data sharing and reproducibility in air pollution epidemiology are of paramount importance, yet the scientific community lacks tools to make this possible (Aim 4).

This report provides a summary of major findings and key conclusions supported by our HEI-funded project.

STUDY AIMS

To overcome these challenges, our team structured our work around four specific aims:

Aim 1: Exposure Prediction and Data Linkage. Investigate the health effects of long-term air pollution in areas with sparse monitoring. Apply and extend already developed and evaluated hybrid prediction models that use satellite, land use, emissions, ground monitoring, and weather data in conjunction with chemical transport models to estimate long-term exposures to low levels of ambient $PM_{2.5}$ mass and components, as well as the gaseous air pollutants O_3 and nitrogen dioxide (NO_2), at high spatial resolution (1 km × 1 km) for the continental United States during the period 2000–2012. Link these predictions to the health data. See Figure 1.

Aim 2: Causal Inference Methods for Exposure–Response. Develop a new framework in Bayesian causal inference to estimate the whole ER that is robust to model misspecification for confounding and accounts for exposure error. Specifically, we aimed to develop methods to: (1) estimate a nonlinear ER, while accounting for exposure error; (2) adjust for measured and unmeasured confounders; (3) adjust for confounding in the context of multiple exposures; and (4) detect effect modification when the multiplicity of possible modifiers precludes testing of each one individually. See Figure 2.

Aim 3: Evidence on Adverse Health Effects. Apply methods developed in Aim 2 to estimate health effects associated with long-term exposure to low levels of ambient air pollution for three dynamic U.S. cohorts: Medicare

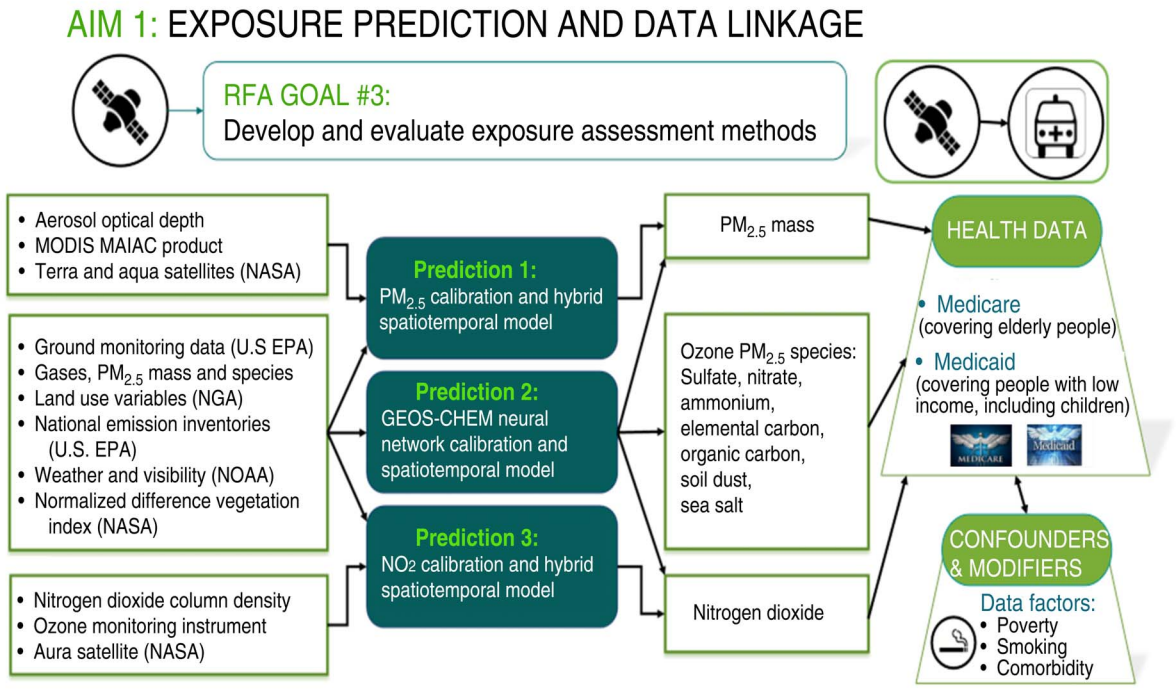


Figure 1. Overview of Aim 1 showing the type, source, and purpose of national data sets used in our prediction models and the corresponding exposure metrics being assessed. (U.S. EPA = U.S. Environmental Protection Agency, NASA = National Aeronautics and Space Administration, NGA = National Geospatial-Intelligence Agency, NOAA = National Oceanic and Atmospheric Administration.)

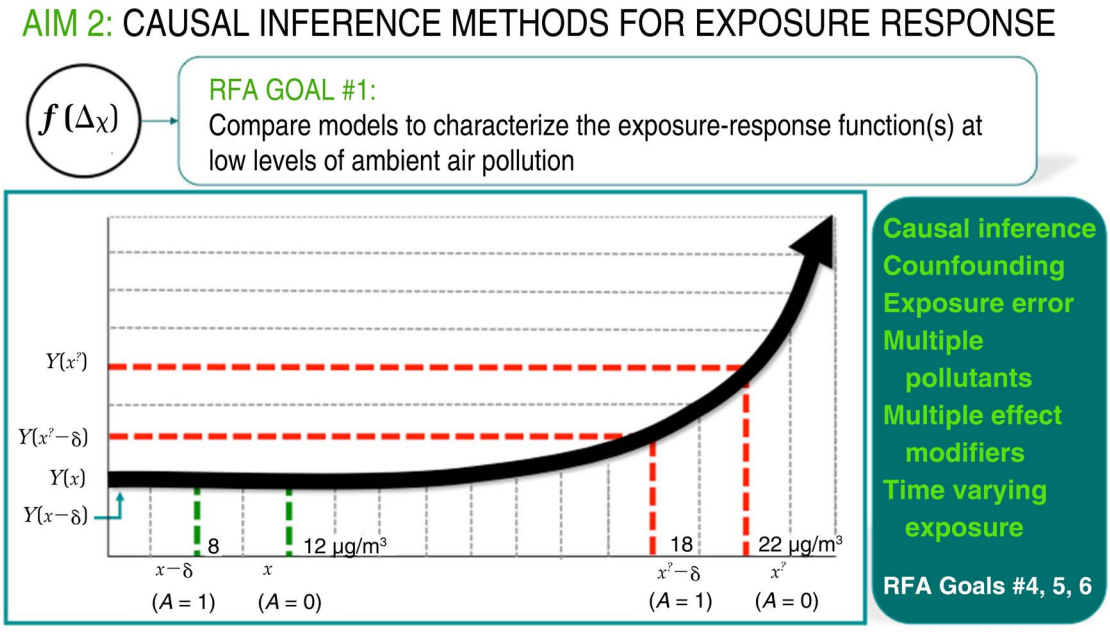


Figure 2. Overview of Aim 2 showing causal inference methods for exposure–response analysis. Causal ER defined as a sequence of hypothetical experiments.

enrollees (28.6 million enrollees [≥65 years] per year, 2000–2012); Medicaid enrollees (28 million enrollees per year, including 12 million children and 7 million people with disabilities, 2010–2012); and MCBS enrollees (nationally representative sample of approximately 15,000 enrollees per year with abundant data on individual-level risk factors, including smoking linked to Medicare claims). We aimed to examine the following health outcomes (time is measured from Medicare enrollment): (1) time to hospitalization by cause; (2) disease progression (time to rehospitalization); and (3) time to death. See Figure 3.

Aim 4: Tools for Data Access and Reproducibility.

Develop tools for reproducible research including approaches for data sharing, record linkage, and statistical software. Figure 4 provides a visual representation of the connection across the proposed aims.

METHODS AND STUDY DESIGN

In Aim 1, we addressed Research Objective #3 of the RFA ("Develop and evaluate exposure assessment methods suitable to estimate exposure to low levels of air pollution at various spatial and temporal scales in large study populations, including populations that reside in areas not covered by routine ground-level monitoring"). We assembled and linked a wealth of data sources from satellite and ground monitoring data and applied, compared, and validated prediction models to estimate long-term average levels of $PM_{2.5}$, $PM_{2.5}$ species, NO_2, and O_3 in a 1 km × 1 km grid for the continental United States. In this aim, we also refined previously developed methods to align gridded exposure to ZIP-code-level exposure including the propagation of the exposure error from grid to ZIP code (location code used by the U.S. Postal Service). We then linked the exposure, health, and confounder data at the ZIP code level (see Figure 1).

In Aim 2, we developed new methods to address Research Objectives #1, 4, 5, and 6 of the RFA ("1. Compare and contrast alternative models and their uncertainty, e.g., threshold/nonthreshold, linear/nonlinear, and parametric/nonparametric, to characterize the ER function(s) at low levels of ambient air pollution. 4. Develop, evaluate, and apply statistical methods to quantify and correct for exposure measurement error in risk estimates and in characterization of ER relationships. 5. Develop and validate approaches to assess the impacts of co-occurring pollutants on health effect associations at low ambient concentrations. 6. Develop and validate indirect approaches to correct risk estimates for the effects of important potential confounding variables, such as smoking, in the absence of such data at the individual level.") We developed several new statistical methods for *causal inference* to estimate the whole ER function. The new methods were designed to overcome several important challenges in the estimation of health effects associated with low-level exposure (see Figure 2).

In Aim 3, we addressed Research Objective #2 of the RFA ("Explore possible variability in effect estimates at low levels among populations, and identify possible contributing factors. Such factors may include age, socioeconomic position, health status, and access to medical care, as well as differences in air pollution sources and time–activity patterns"), providing national evidence on the causal effects of low-level exposure on several outcomes in children, adults with low-income, and adults 65 years and older (see Figure 3).

In Aim 4, we addressed Research Objective #7 of the RFA ("Improve techniques for record linkage and methods for disclosure protection for optimal use of large administrative databases in air pollution and health research"), developing new tools for data access and reproducibility, including statistical software to implement the methods developed in Aim 2 and specific instructions on how to reproduce our analyses (see Figure 4).

This project was approved by the Institutional Review Board of the Harvard T.H. Chan School of Public Health.

STATISTICAL METHODS AND DATA ANALYSIS

Our HEI-funded work comprises an extensive portfolio of analysis and the development of robust statistical methods that can be used to understand the health effects of long- and short-term exposure to low levels of ambient air pollution. In this section, we provide a high-level overview of this work, as grouped into the following four areas: (1) Exposure assessment and data access; (2) Epidemiological studies of ambient exposures to air pollution at low levels; and (3) Methodological contributions in causal inference; and (4) Open science research data platform.

EXPOSURE ASSESSMENT AND DATA ACCESS

PREDICTING AIR POLLUTION: A FLEXIBLE R PACKAGE*

There is strong evidence that ambient exposure to $PM_{2.5}$ increases risk of mortality and hospitalization. Large-scale epidemiological studies on the health effects of $PM_{2.5}$

* A paper describing development of this method (Sabath et al. 2018) was first presented at the 5th IEEE International Conference on Data Science and Advanced Analytics, 1-4 October 2018 and can be found at *arXiv:1805.11534v2 [stat.ML].*

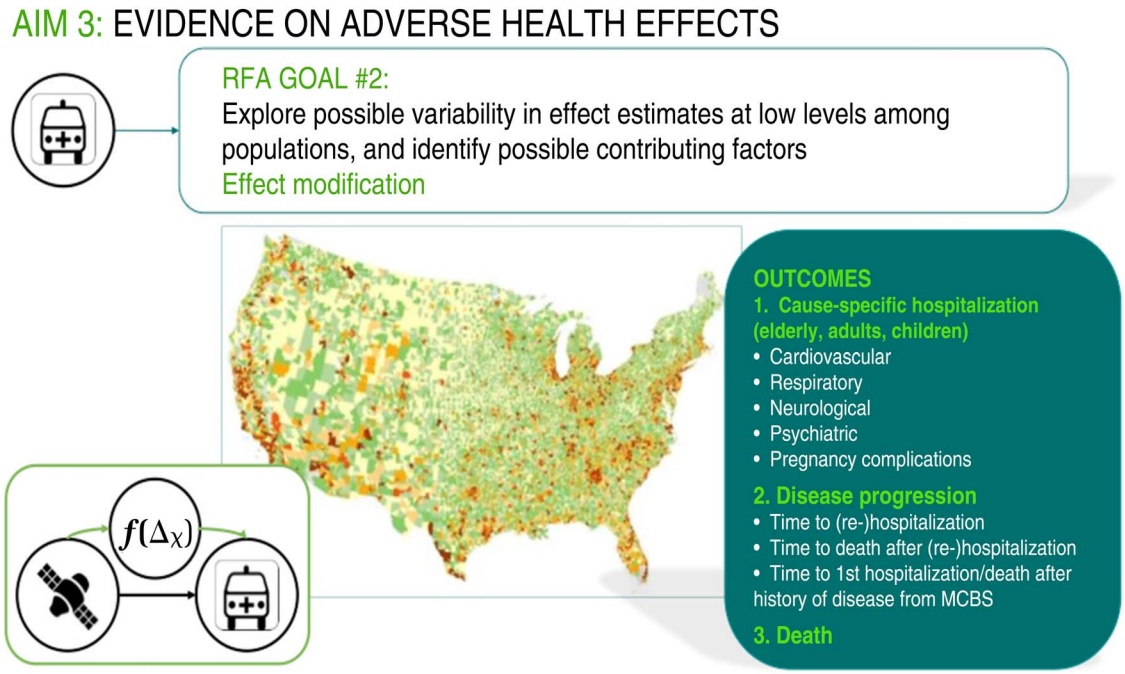


Figure 3. Overview of Aim 3 to provide national evidence on the causal effects of low-level exposure on several outcomes in children, people with disabilities, pregnant women, adults with low-income, and adults 65 years and older.

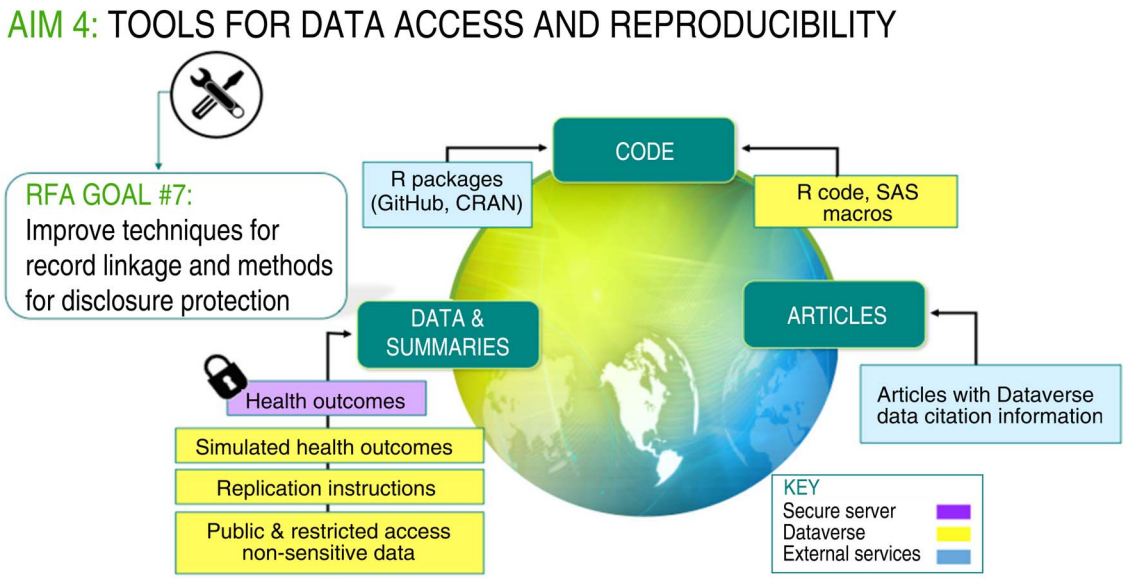


Figure 4. Overview of Aim 4 to develop new tools for data access and reproducibility, including statistical software to implement the methods developed in Aim 2 and specific instructions on how to reproduce our analyses.

provide the necessary evidence base for lowering the safety standards and informing regulatory policy. However, ambient monitors of $PM_{2.5}$ (as well as monitors for other pollutants) are sparsely located across the United States, and therefore studies based only on the levels of $PM_{2.5}$ measured from the monitors would inevitably exclude large parts of the population. One approach to resolving this issue has been developing models to predict local $PM_{2.5}$, NO_2, and O_3 based on satellite, meteorological, and land-use data. This process typically involves developing a prediction model that relies on large amounts of input data and is highly computationally intensive to predict levels of air pollution in unmonitored areas.

Various groups have developed air pollution modeling platforms, with the goal of using information that is available in locations with and without monitors to predict ground level $PM_{2.5}$. A key component used in many of the developed models is satellite-based aerosol optical depth (AOD), a measure of visibility that is associated with levels of particulate matter in the atmosphere. However, these measurements represent particulate matter present in the entire atmospheric column and can't provide a reliable proxy to ground level pollution (Wang et al. 2017). Information from satellite, meteorological and land-use sources is then used to attempt to estimate the levels of $PM_{2.5}$ at surface.

To account for complex atmospheric mechanisms, Di and colleagues (2016; 2017b) used a neural network for its capacity to model nonlinearity and interactions. They employed convolutional layers, which aggregate neighboring information, into a neural network to account for spatial and temporal autocorrelation. We implemented this prediction model to generate daily predictions for the continental United States from 2000 to 2012 for $PM_{2.5}$ and O_3 and tested the performance of the model with monitors left out of the original group (see Figure 5). The model developed by Di and colleagues (2016) relied heavily on data from $PM_{2.5}$ monitors to generate predictions of $PM_{2.5}$. They took a data intensive approach, using information

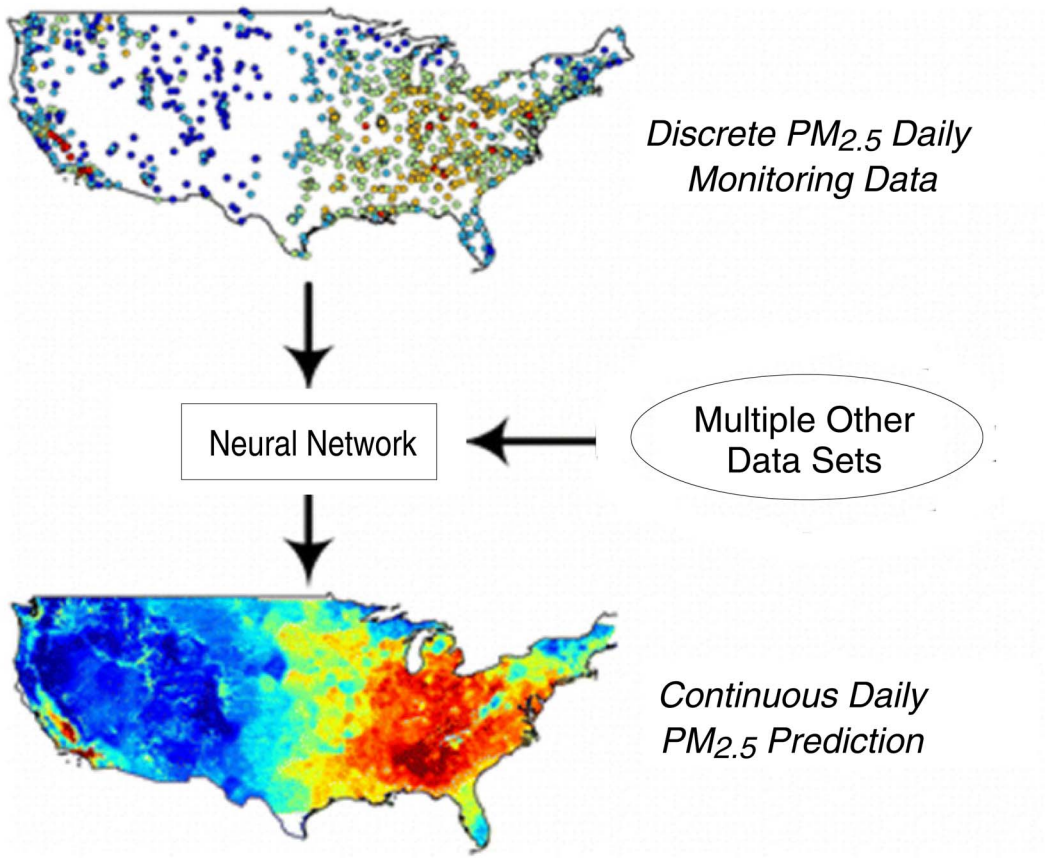


Figure 5. Distribution of pollution monitors and of predictions taken from the model by Di and colleagues (2016). Reprinted with permission from Di et al 2016. Copyright (2016) American Chemical Society.

about AOD from the moderate resolution imaging spectroradiometer (MODIS) satellite, surface reflectance data, estimates of both ground level $PM_{2.5}$ and total levels of aerosols distributed throughout the whole atmospheric column from the Goddard Earth Observing System (GEOS-Chem), meteorological data from the North American Regional Reanalysis project, indices of aerosols that could potentially absorb $PM_{2.5}$, and land-use information such as elevation, road density, vegetation coverage, and population density that can serve as reasonable proxies for emissions as well as help capture small scale variations in PM levels.

One key form of data with universal coverage used to help model PM concentrations and other pollutants is the results of chemical transport models (CTMs). CTMs are computationally intensive atmospheric models that model material flows and chemical reactions within the atmosphere. When combined with measurements of AOD, these have been shown to provide more accurate measurements of $PM_{2.5}$ (Pafka 2015; Wang et al. 2017). Multiple approaches also bring in land-use data to further refine the predictions (Di et al. 2016; Wang et al. 2017). These pieces of information were combined using either standard linear-based statistical models or more complex machine learning methods in order to generate predictions. These are typically able to generate predictions at a 1 km × 1 km scale for large regions, such as the continental United States and the whole of North America (Wang et al. 2017).

This model has been applied in health research to determine the effects of low level $PM_{2.5}$. By incorporating these predictions, this research was able to analyze data from underrepresented populations and demonstrate a connection between increased mortality and PM exposure even at low levels (Di et al. 2017a, c).

A unifying factor among all modeling approaches for estimating exposure prediction — including the model used by Di and colleagues — is that they are both data and computationally intensive. In our work, we used data that took up 30 TBs of disk storage in unprocessed form. The inputs for the $PM_{2.5}$ model came from a variety of sources including atmospheric imaging primarily from National Aeronautics and Space Administration (NASA) satellites, meteorological results of chemical-transport model simulations, geographical information, and information on land usage (such as measurements of road density). For example, inputs for the model by Di and colleagues (2016) are shown in Table 1. Di and colleagues developed their software toolkit in Matlab. It is also worth noting that, because our model was optimized to maximize R^2 with existing monitoring stations, it raises the possibility of overfitting for monitored areas, and this model training process could also make our model less accurate in unmonitored, primarily rural, areas. In order to eliminate the impact of potential overfitting, we trained our models using cross-validation. To address the issue of the model being potentially less accurate in unmonitored areas, we estimated, as a sensitivity analysis, the distribution of population as a function of the distance to the nearest monitoring site and found that 72.9% and 75.1% of the U.S. population lives within 20 km of $PM_{2.5}$ monitoring sites and O_3 monitoring sites, respectively. As a substantial proportion of the U.S. population is located near monitors, our prediction model, which is trained on monitoring stations, would provide accurate estimates for a large proportion of the population.

We have developed a flexible R package called airpred (Sabath et al. 2018) (*https://github.com/NSAPH/airpred*) that allows environmental health researchers to design and train spatio–temporal models capable of predicting multiple pollutants, including $PM_{2.5}$. We utilized the R statistical language together with H2O, an open source big data platform, to achieve both performance and scalability when used in conjunction with cloud or cluster computing systems.

Table 1. Data Sources and Resolutions Used in the Di Model[a]

Input Type[b]	Sources[c]	Spatial Resolution
Meteorological data	Reanalysis	0.5° × 0.625°
AOD	MAIACUS	1 km × 1 km
Surface reflectance	MOD09A1	500 m × 500 m
CTMs	GEOS-Chem CMAQ	12 k × 12 k
Absorbing aerosols	OMAERUVd OMAEROe	0.25° × 0.25°
Vegetation	MOD13A2	1 km × 1 km
Other land use	NLCD	30 m × 30 m

[a] Di et al. 2016.

[b] AOD = aerosol optical depth; CTM = chemical transport model.

[c] MAIACUS (Multi-Angle Implementation of Atmospheric Correction with U.S. data) is a new algorithm to retrieve aerosol optical depth data from a satellite. MOD09A1 is the product name of surface reflectance data from a satellite. GEOS-Chem and CMAQ are chemical transport models. OMAERUVd and OMAEROe are two algorithms for retrieving aerosol composition data. OMAERUVd is a near-ultraviolet algorithm, which retrieves ultraviolet aerosol index, and OMAEROe uses a multiwavelength aerosol algorithm, whose outputs include aerosol indexes at the visible and ultraviolet ranges. MOD13A2 is the data product name of normalized difference vegetation index. NLCD (National Land Cover Database) provides land cover for the entire study area at 30 m × 30 m resolution.

With airpred, we ported and extended the Matlab workflow to the open-source R statistical software (R Development Core Team 2011), while explicitly relying on the H2O "big data" software stack, which is known to perform well on laptops as well as on large computer systems, facilitating the use of parallel algorithms and cluster and cloud deployment options to speed processing. The target audience for the package primarily consists of environmental epidemiology researchers and environmental scientists. As R is a common programming language used in these two fields, we chose to implement the prediction platform in the form of an R package. In addition to user familiarity, R also has the advantage of having a large ecosystem of packages available, allowing for the use of systems developed by others rather than needing to develop many utilities ourselves. Additionally, by developing in R, we can potentially release our package on CRAN, a repository of publicly available R packages that can be accessed via built in functions in R. All CRAN R packages are open source as well, which would allow us to ensure that all code used is publicly available. It was crucial to develop a platform that allows for flexibility, as both the model inputs and the statistical models themselves are frequently changing. Given this, we chose to design a system that would allow users to easily change these elements without making any alterations to the back-end code of the package.

Further, it was important to ensure readability and ease of use for any script utilizing the developed package. Therefore, the number of arguments passed directly to functions was minimized, and the package was designed to include only a small number of clearly named functions that users would need to call in order to implement the full workflow. Airpred can be used to replicate the exposure prediction modeling that provides the air pollution exposure estimates used in the work by Di and colleagues (2017a,c), but users can also specify different types of neural networks, with different parameters, or even perform ensemble modeling.

When modeling $PM_{2.5}$ and other pollutants, we must take into account technical limitations at every step of the process. Moving large quantities of data through memory, let alone feeding them into machine learning (e.g., neural networks) or other modeling systems, can quickly run into system limits. Because of this, it is important to take into account not just the methodology of developing prediction models, but the entire workflow, and to treat all work around it as a single prediction platform. The airpred package implements a single prediction platform for modeling air pollution exposure data. It provides a generic framework to (1) process and assemble raw data sets from a variety of sources; (2) train a deep learning model on the assembled data sets; and (3) generate predictions at the requested spatio–temporal scale. The developed R package is flexible and can be applied to any pollutant. Our R code relies on wrappers to the deep learning algorithms developed in the H2O (*www.h2o.ai/*) open-source software for big data analysis and machine learning at scale.

IMPROVING DATA ACCESS — CREATING AN OPEN SCIENCE RESEARCH DATA PLATFORM

Once we developed the model for estimating exposures to predict $PM_{2.5}$ and O_3 for every 1 km × 1 km grid in the United States (Sabath et al. 2018), we then needed to address data access issues in order to enable investigators to link our curated exposure data to confounder data and health data (from Medicaid and Medicare). Particularly because scientific evidence of harmful effects of air pollution is being subjected to unprecedented scrutiny (Dominici et al. 2014; Samet 2011; Zigler and Dominici 2014), data access and reproducibility are central to current debates on how studies can constitute the scientific base to support regulatory decisions.

To address the need for greater data access and reproducibility, and to achieve Aim 1 and Aim 4 of the project, we posted our data, workflows, and analyses to a secure high-performance computing cluster with the objective of developing an open science research data platform (*https://osf.io/2cg6v/*). Our research data platform contains three distinct but complementary parts:(1) nonhealth data (exposure and confounders); (2) health data (Medicare and Medicaid); and (3) analyses, where data from (1) and (2) are merged and statistical tools are used to address the scientific questions of Aim 3.

Exposure data for $PM_{2.5}$ and O_3 are available at the daily and annual levels at different levels of spatial aggregation (grid, ZIP code, ZIP code tabulation area, county), with different aggregation methods. In Figure 6, we show the $PM_{2.5}$ predictions from Di and colleagues (2016) for New England in the year 2000 aggregated from 1 km × 1 km grids to the ZIP code level, using (a) area-weighting, (b) population-weighting, and (c) inverse-distance weighting. To promote the dissemination of our results, we created several external-facing interfaces that provide interactive visualization of $PM_{2.5}$ and O_3 (e.g., *http://arcg.is/1zTS8S)*. It is important to note that, in order to evaluate the limitations of the data set we used to estimate exposure, we also included in the estimation data from the nearest monitoring sites. While the health effect estimates using air pollution data from proximal monitoring sites were lower than our estimated exposure data, they were still statistically significant. Further, a paper by Wu and colleagues

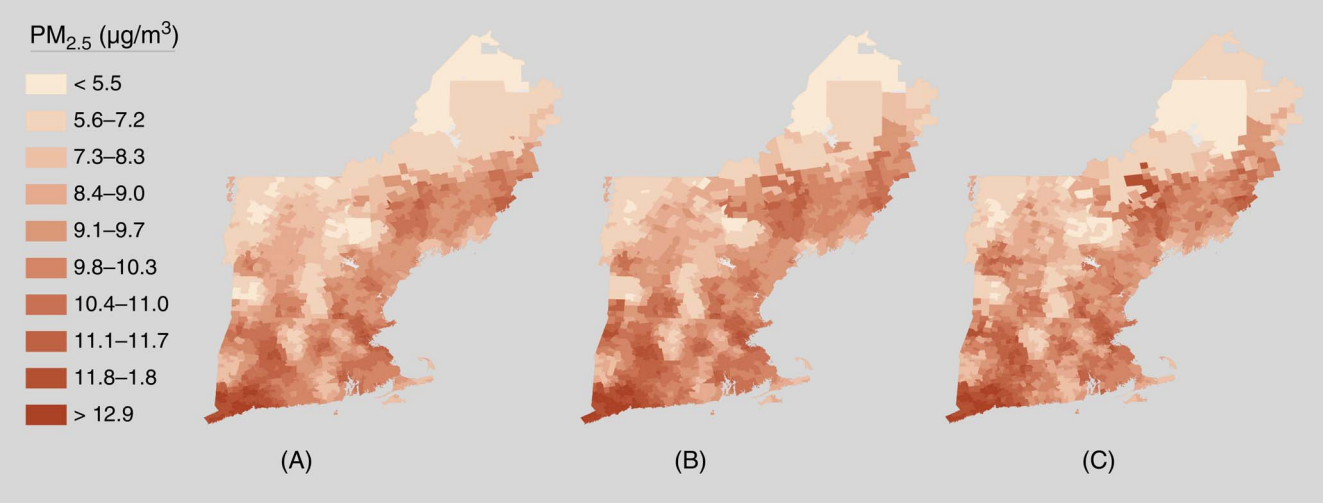


Figure 6. $PM_{2.5}$ predictions for New England in 2000 from the study by Di and colleagues (2016). Predictions were aggregated from 1 km × 1 km grids to the ZIP code level using **(A)** area weighting, **(B)** population weighting, and **(C)** inverse-distance weighting**.**

(2019) deploys novel methods to address the issue of error in air pollution exposure assignment in the context of causal inference. This approach utilizes several different methods to adjust for confounding in causal inference (inverse probability treatment weighting [IPTW] using generalized propensity scores [GPS], GPS matching, and GPS stratification). Using this approach we showed that when adjusting for exposure error, the causal effect of exposure to moderate levels of $PM_{2.5}$ ($8 < PM_{2.5} < 10$ $\mu g/m^3$) causes a 2.8% (95% confidence interval [CI], 0.6%–3.6%) increase in all-cause mortality compared with low exposure ($PM_{2.5} < 8 \mu g/m^3$).

To ensure the reproducibility of our workflow, we developed software codes and packages that allow investigators to link the already curated exposure and confounder data to Medicare and Medicaid claims data. For investigators who own or wish to purchase their own Medicare and Medicaid claims data, we can provide the exact code that we used to link this claims data to the nonhealth data and that we are using to conduct our own analyses of Medicare and Medicaid data. It allows other investigators to reproduce our analyses, replicate our findings, and conduct new analyses. It also guarantees the reproducibility of our own epidemiological analyses that use Medicare and Medicaid claims data to understand the health impacts of environmental exposures. The research data platform is an asset in increasing the scientific rigor of air pollution epidemiological studies by potentially reducing inconsistency of results across studies.

Our code and software tools are under version control. They are hosted on the GitHub social-coding software platform (National Studies on Air Pollution and Health; *https://github.com/NSAPH/airpred*), and we rely on the open science framework (*https://osf.io/2cg6v/*) to provide a searchable web interface to our data and code resources.

It should be noted that the scale of the research data platform in terms of computation and storage resources allows us to undertake studies on the whole Medicare population, such as two studies described in articles we recently published in the *New England Journal of Medicine* (Di et al. 2017c) and in the *Journal of the American Medical Association* (Di et al. 2017a), described later in this report. These two huge studies are *reproducible*: they rely entirely on publicly available data, which are listed in Table 2. In a recent commentary in *Science*, Cosier (2018) pointed to the importance of our work for promoting open, reproducible evidence that can be used to inform public policy.

Table 2. Data Sources[a,b]

Source	Data Set	Website
NOAA	Reanalysis meteorological data	*http://www.noaa.gov/*
NASA	MAIAC AOD data Surface reflectance data NDVI data OMI Aerosol Index Data	*https://www.nasa.gov/*
U.S. Geological Survey	Global terrain elevation data	*https://lta.cr.usgs.gov/*
U.S. Census Bureau	Road density, population count, and area	*https://www.census.gov/*
MRLC	National Land Cover Dataset	*https://www.mrlc.gov/*
GEOS-Chem	Simulation outputs	*http://acmg.seas.harvard.edu/geos/*
U.S. EPA	AQS monitoring data ($PM_{2.5}$ and O_3)	*https://www.epa.gov/aqs*
CMS	Medicare denominator files Medicare Current Beneficiary Survey	*https://www.cms.gov/*
CDC	BMI, smoking rate	*https://www.cdc.gov/*
Dartmouth Atlas of Health Care	Demographics of the Medicare population	*http://www.dartmouthatlas.org/*

[a] A detailed list and software codes are available at *https://osf.io/j6hw8/*.

[b] AOD = aerosol optical depth; AQS = air quality system; BMI = body mass index; CDC = Centers for Disease Control and Prevention; CMS = Center for Medicare and Medicaid Services; GEOS-Chem = Goddard Earth Observing System chemical transport model; MAIAC = Multi-angle implementation of atmospheric correction; MRLC = Multi-Resolution Land Characteristics Consortium; NASA = National Aeronautics and Space Administration; NDVI = Normalized Difference Vegetation Index; NOAA = National Oceanic and Atmospheric Association; OMI = ozone monitoring instrument; U.S. EPA = United States Environmental Protection Agency.

EPIDEMIOLOGICAL STUDIES OF AMBIENT EXPOSURES TO AIR POLLUTION AT LOW LEVELS

LONG-TERM EXPOSURE TO AIR POLLUTION AND MORTALITY IN THE MEDICARE POPULATION*

There is strong evidence that long-term exposure to air pollution leads to increased mortality. Several studies suggest that long-term exposure to $PM_{2.5}$ results in reduced life expectancy; the National Ambient Air Quality Standard is based on such studies. Evidence for mortality resulting from long-term exposure to O_3 is more uncertain, though some studies do suggest reduced survival. However, most air pollution studies have been conducted in urban populations — with higher pollution levels — and evidence is quite limited for any health effects below the standards.

A nationwide cohort study was conducted using all Medicare beneficiaries from 2000 to 2012. Long-term exposure estimates for daily levels of $PM_{2.5}$ and O_3 were developed using hybrid prediction models. We found evidence for statistically significant adverse effects of $PM_{2.5}$ and O_3 exposures at concentrations below current national standards. This effect was greater for self-identified racial minorities and people with low income.

Data and Methods

Mortality among Medicare Beneficiaries. Information about Medicare participants was obtained from the Centers for Medicare and Medicaid Services (CMS) to create a cohort of over 61 million subjects, with a total of over 460 million person-years of follow up. This was an open cohort, where each Medicare subject was followed from the beginning of their first full year of eligibility (age 65 years) during the recruitment period (2000–2012). The outcome of interest was all-cause mortality, and survival

* The following is a summary of an article published in the *New England Journal of Medicine* by Di and colleagues (2017c). A copy of this article, along with a supplementary appendix, is available in Additional Materials on the HEI website, with permission of the publisher.

time was measured from the year subjects entered the cohort until the year of their death (Figure 3).

$PM_{2.5}$ and O_3 Exposure Assessment. Long-term exposure estimates were developed using estimated daily levels of $PM_{2.5}$ and O_3 at 1 km × 1 km resolution for the continental United States during the period 2000–2012 using hybrid prediction models that use satellite, land use, emissions, ground monitoring, and weather data in conjunction with chemical transport models (see Di et al. 2017c, Figure 1). Annual PM and warm season O_3 levels were assigned to each person in the cohort for each calendar year of their inclusion.

Statistical Analysis

Survival analyses were conducted using a variant of the Cox Proportional Hazards Model, known as the Andersen Gill Model (AG) (Andersen and Gill 1982), which allows for the inclusion of covariates that change from year to year.* The results from this model were used to estimate the hazard ratio associated with a 10-μg/m^3 increase in $PM_{2.5}$ and a 10-ppb increase in O_3 exposure. In addition, a surface was fitted to describe the combined and potentially synergistic effects of both $PM_{2.5}$ and O_3 on mortality using a log-linear model with a thin-plate spline. To investigate the effects of lower exposures, separate analyses were conducted that included only person-years with $PM_{2.5}$ exposures lower than 12 μg/m^3 and O_3 exposures lower than 50 ppb ("low-exposure analyses"). A series of subgroup analyses was also conducted to consider the potential interaction of pollutants with key covariates.

Covariate Information

A total of 20 covariates were included in the study, including individual-level covariates, county-level variables, ZIP-code-level variables, three hospital service area–level variables, meteorological variables, and one dummy variable (see Di et al. 2017c, Table 1 and supplementary appendix). Table 3 in this report summarizes the characteristics of the cohort with respect to all of these covariates and provides average levels to which each covariate was exposed in both high and low pollutant regions (for additional information, see supplementary appendix for Di et al. 2017c in Additional Materials on the HEI website). Some of the covariate information was obtained from the MCBS and the Behavioral Risk Factor Surveillance System.

* Note that the use of the AG model is not explicitly discussed in the paper published by Di et al. 2017c.

Results

The Medicare cohort, with more than 61 million individuals and more than 22 million deaths, had excellent power to estimate the risk of death from air pollution over a range of exposure levels, including those below the current NAAQS (Figure 7). In two-pollutant analyses, a 10-μg/m^3 increase in $PM_{2.5}$ was found to be associated with a 7.3% (95% CI, 7.1%–7.5%) increase in mortality for a given ZIP code, and a 10-ppb increase in O_3 concentration was associated with a 1.1% increase (95% CI, 1.0%–1.2%) (Table 4). When the analysis was restricted to person-years with exposure to $PM_{2.5}$ of less than 12 μg/m^3 and O_3 of less than 50 ppb, the same increases in $PM_{2.5}$ and O_3 were associated with increases in the risk of death of 13.6% (95% CI, 13.1%–14.1%) and 1.0% (95% CI, 0.9%–1.1%), respectively.

The subgroup analyses described effect modification for a range of variables (see Di et al. 2017c, Figure 2). Specifically, we found an increase in mortality from exposure to $PM_{2.5}$ among male, black, Asian, and Hispanic subgroups. Stratification by Medicaid eligibility (a measure of socioeconomic status) showed a slightly higher estimated risk from $PM_{2.5}$ exposure than in the general population. The effect of O_3 exposure on mortality was higher among whites and those eligible for Medicaid, but the risk was also below 1 for certain racial subgroups, suggesting nonlinear interaction effects.

Discussion

Using an open cohort of all Medicare participants representing more than 96% of the population of older adults in the United States, our survival analysis demonstrated associations between mortality and long-term exposure to $PM_{2.5}$ and O_3, even at levels below the NAAQS for $PM_{2.5}$ and O_3. Black and Hispanic individuals had a higher risk of death associated with exposure to $PM_{2.5}$ than other subgroups. These findings suggest that lowering the annual NAAQS for $PM_{2.5}$ will produce important public health benefits overall, especially among self-identified racial minorities and people with low income.

Robustness of Risk Estimates. Mortality is influenced by individual-level behavioral risk factors, such as smoking and income, but these data are not included in Medicare claims. To explore the potential impact of such factors, we analyzed the MCBS subsample. The results revealed that the lack of such information did not lead to biased results (see supplementary appendix for Di et al. 2017c in Additional Materials on the HEI website). Our results were also robust to choice of statistical method; we obtained similar risk estimates when individual and

Table 3. Baseline Characteristics of the Entire Cohort

		O_3 Concentration[b]		$PM_{2.5}$ Concentration[b]	
Variable[a]	Entire Cohort	≥50 ppb	<50 ppb	≥12 μg/m^3	<12 μg/m^3
Number of individuals	60,925,443	14,405,094	46,520,349	28,145,493	32,779,950
Number of deaths	22,567,924	5,097,796	17,470,128	10,659,036	11,908,888
Total person-years[c]	460,310,521	106,478,685	353,831,836	212,628,154	247,682,367
Median follow-up years	7	7	7	7	7
Air Pollutants[d]					
Average O_3 (ppb)	46.3	52.8	44.4	48.0	45.3
Average $PM_{2.5}$ (μg/m^3)	11.0	10.9	11.0	13.3	9.6
Individual Covariates[d]					
Male (%)	44.0	44.3	43.8	43.1	44.7
White (%)	85.4	86.6	85.1	82.0	88.4
Black (%)	8.7	7.2	9.2	12.0	5.9
Asian (%)	1.8	1.8	1.8	2.1	1.6
Hispanic (%)	1.9	2.0	1.9	1.9	1.9
Native American (%)	0.3	0.6	0.3	0.1	0.6
Medicaid Eligible (%)	16.5	15.3	16.8	17.8	15.3
Age at entry	70.1	69.7	70.2	70.1	70.0
Ecological Variables[d]					
BMI (kg/m^2)	28.2	27.9	28.4	28.0	28.4
Ever smoker (%)	46.0	44.9	46.2	45.8	46.0
Hispanic population (%)	9.5	13.4	8.4	8.4	10.0
Black population (%)	8.8	7.2	9.3	13.3	6.3
Median household income[e]	47.4	51.0	46.4	47.3	47.4
Median value of housing[e]	160.5	175.8	156.3	161.7	159.8
Below poverty level (%)	12.2	11.4	12.4	12.5	12.0
Below high school education (%)	32.3	30.7	32.7	35.3	30.6
Of owner occupied housing (%)	71.5	71.3	71.6	68.6	73.2
Population density (individual/km^2)	3.2	0.7	3.8	4.8	2.2
With LDL-C test (%)	92.2	92.0	92.2	92.2	92.2
With hemoglobin A1c test (%)	94.8	94.6	94.8	94.8	94.8
With ≥1 ambulatory visit (%)	91.7	92.2	91.6	91.7	91.7
Meteorological Variables[d]					
Temperature (°C)	14.0	14.9	13.8	14.5	13.7
Relative humidity (%)	71.1	60.8	73.9	73.7	69.6

[a] BMI = body mass index; LDL-C = Low-density lipoprotein-cholesterol.

[b] We calculated these summary statistics separately for individuals residing in ZIP codes with average O_3 levels below and above 50 ppb and with $PM_{2.5}$ levels below and above 12 μg/m^3. The value 12 μg/m^3 was chosen as the current annual National Ambient Air Quality Standard (NAAQS).

[c] Total person-years of follow-up in the cohort from 2000 to 2012.

[d] Average values for air pollution levels, ecological variables, and meteorological variables were computed by averaging values over all ZIP codes from 2000 to 2012.

[e] Numbers are presented in U.S. dollars (thousands).

Data from Di et al. 2017c.

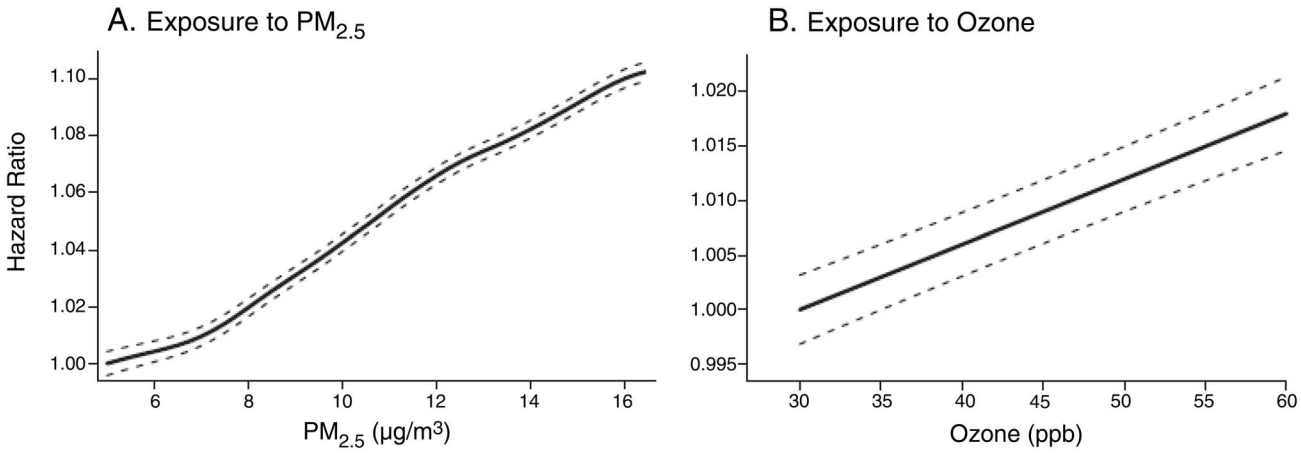


Figure 7. Concentration–response function of the joint effects of exposure to $PM_{2.5}$ and O_3 on all-cause mortality. (Reprinted with permission from Di et al. 2017c, © 2017 Massachusetts Medical Society.)

Table 4. Risk of Death Associated with a 10-μg/m^3 Increase in $PM_{2.5}$ or 10-ppb Increase in O_3 Exposure[a,b]

Model	$PM_{2.5}$ HR (95% CI)	O_3 HR (95% CI)
Two-pollutant analysis		
Main analysis	1.073 (1.071–1.075)	1.011 (1.010–1.012)
Low-exposure analysis	1.136 (1.131–1.141)	1.010 (1.009–1.011)
Nearest-monitor analysis[c]	1.061 (1.059–1.063)	1.001 (1.000–1.002)
Single-pollutant analysis[d]	1.084 (1.081–1.086)	1.023 (1.022–1.024)

[a] Reprinted with permission from Di et al. 2017c, © 2017 Massachusetts Medical Society.

[b] Hazard ratios (95% confidence intervals) for a 10-μg/m^3 increase in $PM_{2.5}$ and a 10-ppb increase in O_3 exposure.

[c] Daily average $PM_{2.5}$ and daily O_3 monitoring data were retrieved from the U.S. EPA Air Quality System (AQS). Daily O_3 concentrations were averaged from April 1 to September 30 to compute warm-season averages. Individuals were assigned to $PM_{2.5}$ and O_3 levels from the nearest monitoring site within 50 kilometers. If there was more than one monitoring site, the nearest one was chosen. Individuals who lived ≥ 50 kilometers away from any monitoring site were excluded.

[d] For the single-pollutant analysis, model specifications were the same as in the main analysis, except that O_3 was not included in the model when estimating the main effect of $PM_{2.5}$ and vice versa.

ecological covariates were excluded from the main analysis, the age of entry was categorized more finely, the estimation procedure was varied (generalized estimating equation [GEE] vs. mixed effects), or a different statistical software was used (R vs. SAS).

Health Effects below NAAQS. Our analyses found a significant association between $PM_{2.5}$ exposure and mortality at concentrations below 12 μg/m^3, with a steeper slope below that level (Figure 7); this suggests that the health benefit from per unit decreases in $PM_{2.5}$ levels is larger for $PM_{2.5}$ concentrations below the current annual NAAQS than for those above that level. Similar steeper concentration–response curves at low concentrations have been observed in previous studies. Significantly, our analyses (down to about 5 μg/m^3) do not provide evidence for a threshold value — the concentration below which $PM_{2.5}$ exposure does not impact mortality.

The O_3 standard in the United States is 70 ppb (daily maximum 8-hour average); there is no annual or seasonal standard. Our analyses, based on annual summer average O_3 concentrations, found an association with O_3 levels, lending support for the establishment of a seasonal or annual O_3 standard. We observed larger effect sizes for longer-term O_3 exposure, even in locations where O_3 concentrations never exceed 70 ppb. The concentration–response curve for O_3, using a two-pollutant model, was linear. Finally, the longer-term all-cause mortality results from this study and similar results from other studies (Turner et al. 2016), suggest that the current O_3 NAAQS — based only on acute effects from daily respiratory mortality — may underestimate the total health burden of effects from O_3 exposure.

Subgroup Analyses. Given the very large sample size, this study afforded an estimation of mortality risk associated with long-term exposure to $PM_{2.5}$ and O_3 among predefined subgroups, such as racial minorities and disadvantaged subgroups. We found that the estimated effect size was larger among male, black, and Medicaid-eligible individuals. To determine whether the effect modifications by race and by Medicaid status were independent, we estimated effects in a subgroup of Medicaid-*eligible* whites and in a subgroup of Medicaid-*ineligible* blacks. We found that blacks not eligible for Medicaid continued to have higher $PM_{2.5}$ associated mortality.

Finally, the $PM_{2.5}$ health effect exhibited an urban–rural difference, which may be due to compositional differences (see Table S3 in the supplementary appendix for Di et al. 2017c available in Additional Materials on the HEI website).

Consistency with Previous Results. The 7.3% increase in risk of all-cause mortality observed in this study for each 10-μg/m^3 increase in $PM_{2.5}$ exposure is within the range of the 13% to 14% increase reported in the Harvard Six Cities Study (Dockery et al. 1993), its extended follow-up study (Krewski et al. 2000), and in other studies. An increased (13.6%) risk for mortality was observed at lower $PM_{2.5}$ concentrations, as reported in the Canadian Community Health Cohort (see Di et al. 2017c for references).

A 10-ppb change in O_3 exposure was associated with a 1.1% increase in the risk of all-cause mortality; this result is close to the 2% increase in all-cause mortality reported by other investigators. This study provides substantial additional evidence that long-term O_3 exposure is associated with a shortened lifespan.

Limitations. Although our exposure models had very good out-of-sample predictive power, like all epidemiology studies of this nature, we cannot fully rule out exposure assessment error issues. Such errors can attenuate the effect estimates in air pollution studies (Spiegelman 2016).

The potential for measured and unmeasured confounding bias is an intrinsic concern for all observational studies on air pollution and health. In order to mitigate these concerns, we estimated the low-exposure effects using two distinctly different study designs: (1) the AG model (Andersen and Gill 1982) for estimating the effect of long-term exposure to air pollution on mortality, as described in the study by Di and colleagues (2017c); and (2) a case–crossover study model for estimating the effects of short-term exposure to air pollution and mortality (see next section). These two study designs are subject to different types of both measured and unmeasured confounding bias, which we discuss in further detail in the Conclusions section of this report.

We have examined the effects of only two pollutants in this study. Our reasons were that these are both critical criteria pollutants and that their national standards — NAAQS — are currently scheduled for revision. We recently developed ensemble exposure assessment models for both $PM_{2.5}$ and NO_2 and updated our O_3 model; we expect to include these results in the HEI final report.

Some of the results presented in Figure 2 of the article by Di and colleagues (2017c) for effect modification are puzzling. For example, the protective effect observed for Native Americans, Hispanics, and Asians does not seem biologically based, nor is there an explanation for the difference in hazard ratios for males and females. Given the limits of the regression model for confounding, it is not possible to discern whether these differences may be attributable to model misspecification and/or confounding bias. For these and other reasons, we are developing new methods for causal inference that will give us greater confidence in the results. We plan to apply these new methods to analyze the data so that we may better understand the true effects of ambient air pollution.

ASSOCIATION OF SHORT-TERM EXPOSURE TO AIR POLLUTION WITH MORTALITY IN OLDER ADULTS*

Introduction

The evidence for the health effects of short-term exposure to $PM_{2.5}$ and O_3 is provided by many studies, though these studies have primarily been conducted in populations living in large, well-monitored urban areas, with relatively high levels of pollutants. The study we describe

* The following is a summary of an article published in the *Journal of the American Medical Association* by Di and colleagues (2017a), which may be viewed at *https://jamanetwork.com/journals/jama/fullarticle/2667069* (open access; courtesy of JAMA).

below was designed to study the health effects of exposure to $PM_{2.5}$ and O_3 that included groups living in unmonitored, rural areas. We also sought to shed light on short-term health risks among specific subgroups (e.g., underrepresented minorities and those with low socioeconomic status, such as persons eligible for Medicaid).

We have conducted a case–crossover study to investigate all deaths among Medicaid enrollees in the United States during the years 2000 to 2012; our goal was to estimate mortality risk associated with short-term exposures to $PM_{2.5}$ and O_3 in the general population and in specific subgroups. We also wanted to estimate the association between daily mortality and exposure levels below the current NAAQS.

Methods

Population, Health and Covariate Data. All deaths among all Medicare beneficiaries were abstracted from Medicare claims data for the period 2000 to 2012 along with identifying data, relevant covariates, and the date of death. Individuals with an unverified date of death or still living after December 31, 2012, were excluded. A total of 22,433,862 deaths were identified.

Since confounders are, by definition, correlated with exposure, the only covariates included in the model were those which co-vary with daily air pollution levels. These included air and dew point temperatures along with the alternate air pollutant (O_3 in the case of $PM_{2.5}$ and $PM_{2.5}$ in the case of O_3).

Individual-level and ZIP-code-level covariates that could exhibit little or no change over the course of a month (e.g., age, sex, race/ethnicity, socioeconomic status, smoking, and other behavioral risk factors) were not considered to be potential confounders. However, because these variables may be relevant as effect modifiers, we abstracted age, sex, race, ethnicity, and eligibility for Medicaid from the Medicare and Medicaid records to assess the associations of mortality with $PM_{2.5}$ and O_3 concentrations in potentially vulnerable subgroups.

Exposure: Ambient $PM_{2.5}$ and O_3 at the ZIP Code Level. Daily ambient levels of $PM_{2.5}$ (24-hour average) and O_3 (8-hour maximum in warm season) for all United States at the ZIP code level were estimated using the methods described above. For each subject, ambient concentrations were determined for the "case day," defined as the date of death, and its control days at the ZIP code of residence. For the same person, 3 or 4 control days were chosen and included in the analyses.

Analysis. By fitting a conditional logistic regression to all pairs of case days and matched control days, we estimated the relative risk (RR) of all-cause mortality associated with short-term $PM_{2.5}$ and O_3 exposure (pollutant levels were adjusted for each other). Potential for residual weather-related confounding was controlled by using natural splines of air and dew point temperatures with three degrees of freedom. For each case day, daily exposure to air pollution was defined as the mean of the same day of death (lag 0-day) and 1 day prior (lag 01-day). Relative risk increase (RRI) was defined as RR – 1.

We performed subgroup analyses by sex, race, or ethnicity (white, nonwhite, and others), age brackets (≤69, 70–74, 75–84, and ≥85 years), eligibility for Medicaid, and population density at residence (in quartiles). We fitted separate conditional logistic regressions to the data for each subgroup and obtained subgroup-specific estimates of RR and absolute risk difference. Subgroups were compared using a two-sample test for assessing statistically significant differences in the estimated RR between categories.

In order to focus on effects below the current standards, subanalyses were conducted with cases restricted to those occurring on days with daily air pollution concentrations below 25 $\mu g/m^3$ for $PM_{2.5}$ and 60 ppb for O_3. This reduced the total number of cases to 20,955,387, a 6.6% decrease. ER curves were estimated between $PM_{2.5}$ or O_3 and mortality by replacing linear terms for the two pollutants with penalized splines for both $PM_{2.5}$ and O_3.

Results

These analyses included more than 22 million days with deaths and more than 76 million control days among Medicare enrollees. We found an increase in the daily risk of mortality of 1.05% (95% CI, 0.95%–1.15%) for a 10-$\mu g/m^3$ increase in $PM_{2.5}$ and 0.51% (95% CI, 0.41%–0.61%) for a 10-ppb increase in O_3 among older Americans in the two-pollutant model (Table 5). These associations remained significant when examining days below 25 $\mu g/m^3$ for $PM_{2.5}$ and below 60 ppb for O_3, but with larger effect size estimates, particularly for both $PM_{2.5}$ and O_3 (RRI:1.61% [95% CI, 1.48%–1.74%] and 0.58% [95% CI, 0.46%–0.70%] respectively).

The estimated ER relationships are shown in Figure 8 as the RRI associated with a 10-$\mu g/m^3$ increase in $PM_{2.5}$ or a 10-ppb increase in O_3. Note that the increase in RRI rises sharply for both pollutants at a relatively low concentration — consistent with low-exposure analysis — and then levels out. Importantly, neither curve indicates a threshold for mortality at low concentrations.

In subgroup analyses, significant effect modification was observed for some key covariates (Di et al. 2017a, Figures 3 and 4). Thus, higher mortality was observed among

Table 5. Relative Risk Increase of Daily Mortality Associated with Each 10-μg/m^3 Increase in $PM_{2.5}$ or Each 10-ppb Increase in O_3[a]

Model	Relative Risk Increase% (95% CI)	
	$PM_{2.5}$	O_3[b]
Two-pollutant analysis		
Main analysis[c]	1.05 (0.95–1.15)	0.51 (0.41 –0.61)
Low-exposure[d]	1.61 (1.48–1.74)	0.58 (0.46–0.70)
Nearest monitors[e]	0.83 (0.73–0.93)	0.35 (0.28–0.41)
Single-pollutant[f]	1.18 (1.09–1.28)	0.55 (0.48–0.62)

$PM_{2.5}$ = fine particulate matter; ppb = parts per billion

[a] Adapted with permission from Di et al. 2017a, © 2017 American Medical Association. All rights reserved.

[b] O_3 analyses included days from the warm season only (April 1 to September 30).

[c] The main analysis used the mean of daily exposure on the same day of death and 1 day prior (lag 01-day) as the exposure metric for both $PM_{2.5}$ and O_3, and controlled for natural splines of air and dew point temperatures with 3 df. The main analysis considered the two pollutants jointly included in the regression model and estimated the percentage increase in the daily mortality rate associated with a 10-μg/m^3 increase in $PM_{2.5}$ exposure adjusted for O_3 and the percentage increase in daily mortality rate associated with a 10-ppb increase in warm-season O_3 exposure adjusted for $PM_{2.5}$.

[d] The low-exposure analysis had the same model specifications as the two-pollutant analysis and was constrained for days when $PM_{2.5}$ was below 25 μg/m^3 or O_3 was below 60 ppb.

[e] $PM_{2.5}$ and O_3 monitoring data were retrieved from the U.S. Environmental Protection Agency Air Quality System, which provides the daily mean of $PM_{2.5}$ and daily 8-hour maximum O_3 levels at each monitoring site. Daily O_3 concentrations were averaged from April 1 to September 30. Individuals were assigned to the $PM_{2.5}$ and O_3 levels from the nearest monitor site within 50 km. Those living ≥ 50 km from any monitoring site were excluded.

[f] The single-pollutant analysis estimated the percentage increase in the daily mortality rate associated with a 10-μg/m^3 increase in $PM_{2.5}$ exposure without adjusting for O_3 and the percentage increase in the daily mortality rate associated with a 10-ppb increase in O_3 exposure without adjusting for $PM_{2.5}$.

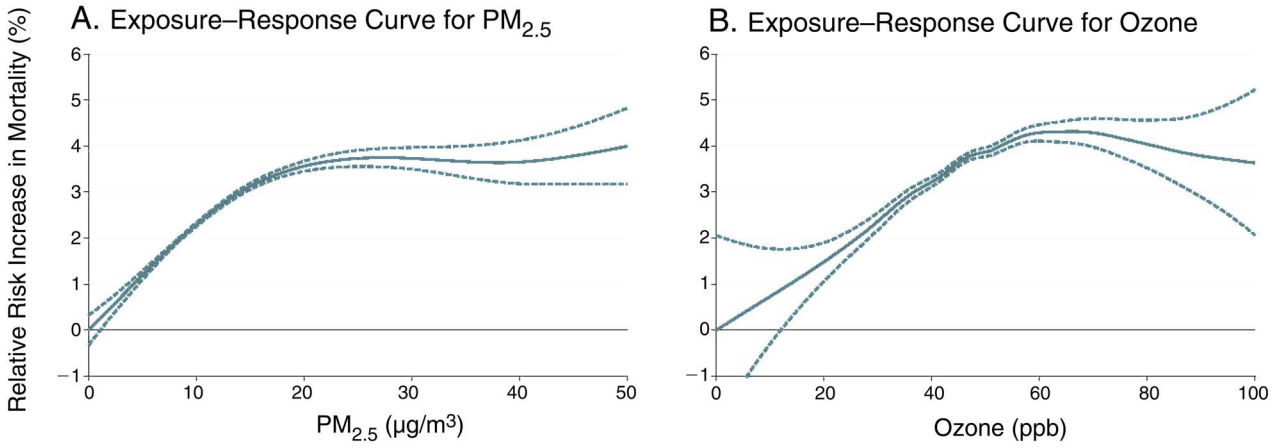


Figure 8. Estimated exposure–response curves for short-term exposures to $PM_{2.5}$ and O_3 (Di et al. 2017a). A two-pollutant analysis with separate penalized splines on $PM_{2.5}$ (A) and ozone (B) was conducted to assess the percentage increase in daily mortality at various pollution levels. Dashed lines indicate 95% CIs. The mean of daily exposure on the same day of death and 1 day prior (lag 01-day) were used as metrics of exposure to $PM_{2.5}$ and ozone. Analysis for ozone was restricted to the warm season (April to September). (ppb = parts per billion.) (Reprinted with permission from Di et al. 2017a, © 2017 American Medical Association. All rights reserved.)

those eligible for Medicaid (RRI: 1.49% [95% CI, 1.29%–1.70%] $P < 0.001$), those older than 70 years (e.g., for ≥85 years, RRI: 1.38% [95% CI, 1.23%–1.54%, interaction: $P < 0.001$), and among females (RRI: 1.20% [95% CI, 1.07%–1.33%], interaction: $P = 0.02$). The effect estimates for $PM_{2.5}$ increased with age and were higher for black individuals than for white individuals ($P = 0.001$). Similar patterns were observed for O_3, but with less contrast between groups.

Discussion and Conclusion

In this large case–crossover study of all Medicare-recipient deaths between 2000 and 2012, both $PM_{2.5}$ and warm-season O_3 exposures were associated with statistically significant increases in mortality. When the analyses were restricted to days with $PM_{2.5}$ and O_3 levels lower than the current daily NAAQS levels, the risk of mortality remained statistically significant. Since Medicaid represents a national cohort, it included people living in smaller cities and rural areas that were unmonitored and thus excluded from previous time-series studies; we observed no difference in mortality risks among urban versus rural residents. This study provides evidence that short-term exposures to $PM_{2.5}$ and O_3 are associated with increased all-cause mortality, even at levels well below the current daily standards. Certain groups — such as older people, females, and those with lower income — are at an elevated risk.

The results of our study are consistent with the results of a large number of previous studies, conducted on cohorts in both the United States and Europe. Our results showing health effects after exposure to levels below the current NAAQS standards are significant, as are the results showing higher increases in mortality rates in certain subgroups — such as Medicaid-eligible individuals, females, and older individuals.

The strengths of this study are that it is based on the largest cohort among all time-series studies of $PM_{2.5}$ and O_3 exposure and health effects; it used state-of-the-art daily exposure assessment techniques for both monitored and unmonitored areas; it had sufficient statistical power to analyze mortality among potentially vulnerable subgroups; and, finally, that it used the case–crossover design, which controls for many confounding factors. Limitations of the study include that the Medicare population comprises individuals 65 years and older, Medicare files do not have information on cause-specific mortality, and there may be some exposure assessment error because exposure was estimated by residential ZIP code and not the exact location of death.

In conclusion, we report that daily $PM_{2.5}$ and warm season O_3 levels are associated with a risk of increased mortality, and this risk was observed at levels below the current standards for the two pollutants.

EXAMINING CAUSAL INFERENCE BETWEEN AIR POLLUTION AND MORTALITY IN THE CONTEXT OF AN ERROR-PRONE EXPOSURE*

Observational studies to estimate the effects of exposure are well-known to be susceptible to sources of bias, particularly exposure measurement error and confounding. We have developed a new approach for estimating causal effects in the presence of exposure error; confounding is adjusted using a GPS. Monitoring data, assumed to be error free, were used as validation data. We then employed a regression calibration (RC)-based adjustment for continuous data for exposure (error prone) and combined it with GPS, thereby adjusting for confounding (RC-GPS). After transforming the corrected continuous exposure into a categorical exposure, we conducted the outcome analysis. We also considered confounding adjustment in the context of GPS subclassification, IPTW, and matching. We found that in simulations with varying degrees of exposure error and confounding bias, as compared with standard approaches that rely on the error-prone exposure, RC-GPS eliminates bias from exposure error and confounding.

To test this approach, we estimated the causal effect of long-term exposure to $PM_{2.5}$ on mortality in New England states for the period from 2000 to 2012 by applying RC-GPS to a rich data platform. We included 2,202 ZIP codes in the main study, covered by 217,660 1 km × 1 km grid cells with yearly mortality rates, yearly $PM_{2.5}$ averages estimated from a spatio–temporal model (error-prone exposure), and several potential confounders. For internal validation, we included a subset of 83 1 km × 1 km grid cells within 75 ZIP codes from the main study with error-free yearly $PM_{2.5}$ exposure data obtained from monitor stations. Under assumptions of noninterference and weak unconfoundedness, we found that exposure to moderate levels of $PM_{2.5}$ ($8 < PM_{2.5} < 10$ μg/m^3) causes a 2.8% (95% CI, 0.6%–3.6%) increase in all-cause mortality compared with low exposure ($PM_{2.5} < 8$ μg/m^3). (See Wu et al. 2019 for additional details.)

We were also interested in exploring the relationship between the grid of estimated ambient concentration and individual exposures. In this report, we have used ZIP-code-level ambient concentration as a proxy for actual individual exposure. To investigate this, we further downscaled our

* A full description of this study was published in the *Annals of Applied Statistics* (Wu et al. 2019).

updated $PM_{2.5}$ model, from 1 km × km to 100 m × 100 m, with additional downscaling of land-use variables. The refined 100-m^2 predictions are quite close to the individual-level exposure estimation that uses the subject's home address. Unfortunately, the Medicare data include only the participant's residential ZIP code.

ADDRESSING LOCAL CONFOUNDING IN EXPOSURE–RESPONSE ESTIMATION*

To address the issue of confounding, we have developed a Bayesian framework for the estimation of a causal ER curve. This framework, called local ER confounding adjustment (LERCA), allows for different confounders and different strengths of confounding at different exposure levels. It also takes into account model uncertainty regarding confounders' selection and the shape of ER curve. Finally, LERCA enables systematic evaluation of the observed covariates' confounding importance at different exposure levels.

Using simulation studies, we compared LERCA with several state-of-the-art causal inference approaches for ER estimation. In addition, we applied this method to health, weather, demographic, and pollution data for 5,362 ZIP codes during the years 2011 to 2013. (An R package is available at *https://github.com/gpapadog/LERCA*.)

The LERCA approach to flexible estimation of the ER curve in observational studies is innovative because:

- Within a potential outcome framework, it casts the formulation of the ER (and several randomized experiments have been mimicked across exposure levels);
- The experimental configuration is informed by the data;
- It allows for different sets of covariates and confounders at different exposure levels (as is the case with our data example);
- It allows for varying confounding across different levels of the exposure;
- It increases efficiency by performing local covariate selection especially at low exposure levels;
- In the posterior inference on the whole ER curve, it propagates model uncertainty for the experiment configuration and covariate selection; and
- It provides important scientific guidance in terms of which covariates are confounders at different exposure levels.

* A full description of this study will be available in an article by Papadogeorgou and Dominici, submitted for publication.

In addition to its application to the estimation of the health effects of air pollution, the LERCA framework provides a data-driven approach that is applicable to many regulatory settings addressing the safety of potentially harmful substances. The method could be routinely used to assess health effects of low-level exposures to such pollutants as lead, environmental contaminants, radiation, and pesticides.

DISCUSSION AND CONCLUSIONS

KEY CONTRIBUTIONS

Our body of work advanced by HEI lends extensive evidence that short- and long-term exposure to $PM_{2.5}$ and O_3 is harmful to human health, increasing the risks of hospitalization and death, even at levels that are well below the NAAQS. Specifically, our HEI-funded work has enabled the following scientific contributions:

1. We have conducted the largest study to date on long-term exposure to $PM_{2.5}$ and O_3 (a prospective cohort with a time-varying exposure) for the entire Medicare population, with an assessment of the risk of mortality at exposure levels that are always below the NAAQS for yearly average (Di et al. 2017c).
2. We have completed the largest study to date on short-term exposure to $PM_{2.5}$ and O_3 (a case–crossover study) for the entire Medicare population, with an assessment of the risk of mortality at exposure levels that are always below the NAAQS for yearly average (Di et al. 2017a).
3. We have advanced the development of an artificial neural network (Di et al. 2017a) and open source R package (Sabath et al. 2018) to estimate daily exposures to $PM_{2.5}$ and O_3 for 11 million 1 km × 1 km grids for the continental United States.
4. We have developed new methods for causal inference to propagate the error in the exposure predictions into the health effects estimation using GPSs (Wu et al. 2019).
5. We have developed new methods for causal inference to estimate a causal ER function allowing a different set of confounders at different levels of exposure (*https://arxiv.org/abs/1806.00928*).

STRENGTHS

There are several factors that contribute to the strengths of this work:

Nation-Wide Analyses at an Unprecedented Scale. We conducted two nation-wide cohort studies (Di et al. 2017a, 2017c), allowing us to examine the health effects of air

pollution exposure at a scope, scale, and level of specificity beyond any previous studies. Our studies, which included claims data from the entire U.S. Medicare population from 2000–2012, included over 460 million person-years of follow up. Health data were linked to highly defined air pollution exposure predictions ($PM_{2.5}$ and O_3) for every 1 km × 1 km scale within the United States, allowing our team to examine the long- and short-term effects of air pollution exposure from a unique and unprecedented perspective.

Two Different Study Designs. Each of our nationwide studies uses different study designs. Di and colleagues (2017c) used a prospective cohort design, allowing us to link $PM_{2.5}$ and O_3 exposure data to health outcomes and mortality data derived from Medicare claims (from 2000–2012). In Di and colleagues (2017a), we used a case–crossover design to examine all deaths of Medicare participants in the continental United States from 2000–2012 and to estimate the mortality risk associated with short-term exposures to $PM_{2.5}$ and O_3 in the general population as well as in subgroups. Although we used two different study designs, both studies provided consistent evidence of a significant increase in mortality risks as levels of $PM_{2.5}$ and O_3 exposure increase, even at levels well below the NAAQS.

Numerous Sensitivity Analyses. Because Medicare claims do not include individual-level data on behavioral risk factors (such as smoking and income), which could impact mortality and thus be important confounders, we conducted multiple sensitivity analyses to assess potential confounding bias. Specifically, we used the MCBS (a nationally representative sample of approximately 15,000 Medicare enrollees per year with high-quality data on individual level risk factors, including smoking) to analyze the influence of potential confounding factors (i.e., age of entry in Medicare; body mass index; current smoking status; smoking history; income). We found that these additional individual-level variables, which were available to only part of the Medicare population, had little effect on the mortality and hospitalization risks of exposure to $PM_{2.5}$ for the entire Medicare population. In addition, we found that our results are robust regardless of the statistical methods used, obtaining similar risk estimates when we: (a) excluded individual and ecological covariates from the main analysis (Additional Materials, Table S2, available on the HEI website); (b) finely categorized age at entry (3- and 4-year rather than the 5-year categories used in the main analysis [see Di et al. 2017c, Figure S3]) (c) varied the estimation procedure (GEE vs. mixed effects) (Additional Materials, Di et al. 2017c, Table S3 and S4); and (d) used different statistical software (R vs. SAS) (see Di et al. 2017c Tables S3 and S4). Finally, we found that our results were consistent with other results published in the literature (see Di et al. 2017c, Figure S6).

In addition, we conducted sensitivity analyses of our air pollution prediction model, obtaining $PM_{2.5}$ and O_3 concentration data from the U.S. EPA Air Quality System (AQS) and matching it to each individual, based on the nearest monitoring site within a distance of 50 kilometers (see Additional Materials, section S.1 of the supplementary appendix for details).

Development of New Methods for Causal Inference. In order to address significant methodological gaps limiting the current environmental epidemiological literature, we developed two novel methods. First, we developed a new approach to comprehensively assess the causal impact of exposure error overall by estimating causal effects when the exposure is measured with error and the confounding adjustment is performed via a GPS. Using categorical GPS is a critical advancement over the more commonly used binary propensity score confounding adjustment approaches. Second, we developed a Bayesian framework to estimate a causal ER curve called LERCA to account for: (a) different confounders and different strengths of confounding at the different exposure levels; and (b) model uncertainty regarding the selection of confounders and the shape of the ER curve. Also, LERCA provides a principled way of assessing the observed covariates' confounding importance at different exposure levels.

Serious Effort to Develop Open Source Software and Increase Reproducibility and Transparency. Through our HEI-funded work, we have made enormous strides in advancing a framework and a set of tools needed to support greater data access, transparency, and reproducible research through an open science research data platform. To ensure the reproducibility of our workflow, we developed software codes and packages that allow investigators to link the already curated exposure and confounder data to Medicare and Medicaid claims data. For investigators who own or wish to purchase their own Medicare and Medicaid claims data, we can provide the exact code that we have used to link this claims data to the nonhealth data that we are using to conduct our own analyses of Medicare and Medicaid data. It allows other investigators to reproduce our analyses, replicate our findings, and conduct new analyses. It also guarantees the reproducibility of our own epidemiological analyses that use Medicare and Medicaid claims data to understand the health impacts of environmental exposures. Ultimately, the research data platform we have developed is an asset in increasing the scientific rigor of air pollution epidemiological studies by potentially reducing inconsistency of results across studies.

LIMITATIONS

Measured and Unmeasured Confounding Bias

The possibility of measured and unmeasured confounding bias is real in all observational studies on air pollution and health. To mitigate the potential impacts of measured and unmeasured confounding bias, we purposely estimated the health effects of low exposure to $PM_{2.5}$ and O_3 on mortality using two very distinct study designs: (1) using the AG model for estimating the effect of *long-term* exposure to air pollution on mortality (Di et al. 2017c); (2) using a case–crossover study for estimating the effect of *short-term* exposure to air pollution and mortality (Di et al. 2017a). These two study designs are subject to different types of both measured and unmeasured confounding bias. In the AG model we need to worry about unmeasured time trends and measured and unmeasured confounders that vary spatially. On the other hand, the case–crossover study is not subject to confounding bias by time trends nor by spatially varying covariates because it exploits the day-to-day variation in air pollution and mortality within each location (ZIP code) and because it leverages matching methods. Conducting both of these studies in parallel was an enormous undertaking. It was a serious attempt to see if we could obtain consistent results when analyzing the data in two completely different ways, knowing that each study would be subject to different sources of bias.

In fact, when applied to the same data, these two study designs, which use two totally different statistical analyses, both reported statistically significant associations, including at low levels of exposure. This increases our level of confidence that, overall, short- and long-term exposures to air pollution are both harmful to human health. The AG model, although less than perfect, has the following features compared with a standard Cox model: (1) The Medicare cohort is a dynamic cohort where new enrollees enter into the cohort every year. In the AG model, follow-up years are defined as the number of years since one participant enters the Medicare program until he or she dies or the study ends (in 2012). (2) To account for the dynamic nature of the cohort, the AG model formulation has the advantage of creating different strata of the population for each follow-up time. In other words, in the AG model, we control for follow up by design. This formulation is similar to a log-linear Poisson model with follow up included as a factor term in the model, except it is parametrized differently; the risks for each follow-up year are proportional in the Poisson model, but not in the AG model. However, we do have multiple calendar years of entry and, although related, follow-up year is not identical to calendar year. The AG formulation captures the effects of air pollution from both spatial and temporal variations. The model estimates a single coefficient for both $PM_{2.5}$ and O_3 by combining information across all strata. The overall estimate is therefore a combination of contributions *that could potentially be vulnerable to confounding by time trends and others that are not.* (3) Extensive sets of measured spatial confounders are available in the data sets; however, most temporal confounders are unmeasured and thus impossible to adjust for by direct inclusion into the AG model. Inclusion of a year as a linear term to adjust for unmeasured confounding bias due to the time trends in the AG model is inadequate as it will eliminate all the temporal variation in the data and result in an over adjustment. Since $PM_{2.5}$ declined between 2000 and 2012, and in addition, the trend has high collinearity with year, it is extremely hard to disentangle the effect of $PM_{2.5}$ from time trends based on the current methodology. It is also the reason that we are currently conducting sensitivity analyses using newly developed causal inference approaches.

We are currently conducting sensitivity analyses and have made enormous progress toward increasing the computational efficiency of fitting statistical models for the continental United States population. As detailed later in our planned analyses, we are developing an approach to reanalyze the same data used in our earlier study (Di et al 2017c), but with two important modifications: (1) we adjust for measured confounding bias using a causal inference approach using a GPS model; and (2) we adjust for unmeasured confounding bias, such as time trend, by including time into the GPS model. The results of these sensitivity analyses will be presented at the end of the study period and in the final HEI report.

Need for Further Application of New Causal Inference Methods in National Epidemiological Studies

The development of new methods for causal inference in the context of these massive data sets whose data vary in space and time is a highly complex endeavor. In this report we described two of our studies — by Wu et al. (2019) and Papadogeorgou and Dominici (unpublished results) — that used new methods. We examined the methods' theoretical properties, tested them in simulation studies, and applied them to a subset of the entire U.S. Medicare data set, specifically focusing on the New England region. However, we recognize that these methods have not yet been developed in the context of the same study designs as those used in our earlier studies (Di et al. 2017a,c). Indeed, we are working on extensions to allow us to apply these methods to the exact same study designs as those used in our previous national cohort studies. We will continue to

develop these causal inference methods further for spatio–temporal data and to scale up the computations for the continental United States.

Need for Formal Propagation of the Exposure Error in Health Effects Estimation

These two national studies do not account for a formal propagation of the exposure error into the health effects estimation. Addressing exposure error is a key priority, and we have attempted to address this issue in three ways: (1) validate the accuracy of our predictions; (2) conduct extensive sensitivity analyses that use only data observed from the monitors; and (3) develop new statistical methods that are purposely designed to propagate the error associated with the prediction of air pollution exposure into the estimation of causal effects of air pollution in health. As detailed in Wu and colleagues (2019), this is a complex problem. The complexity of this task is due to the fact that, in the context of causal inference methods, error in the exposure will affect the regression coefficient measuring the health risks of air pollution exposure (as in any measurement error model), but it will also affect the propensity score model used in the causal inference analysis. In the study by Wu and colleagues (2019), we have detailed an innovative and well-validated approach to overcome these challenges. Although our work so far is limited to the setting where the exposure is categorical and has only been applied to the New England region, we are working on extending it to the continuous setting and applying the approach to the continental United States.

Mobility Bias

Both of these national epidemiological studies are potentially impacted by mobility bias and we are working towards a better understanding this issue. We have found that approximately 21% of Medicare recipients changed residential ZIP code at least once between 2000 and 2012. It could thus be possible that older people who are healthy are more likely to move to more rural, less polluted areas; or that older people with health problems may tend to move into cities to have better access to medical care. Since we update exposure level for each participant every year, their moving will not affect our exposure assessment. We are conducting analyses of long-term effects of $PM_{2.5}$ on mortality separately among the movers and the nonmovers. So far, we have been able to link exposure data to 10,679,150 movers and 52,746,548 nonmovers (unpublished results).

NEXT STEPS

Update the Exposure Estimation for PM, O_3, and NO_2 to 2016

We have updated our $PM_{2.5}$ exposure assessment to 2016 using an advanced modeling strategy. For the new model, we fit an ensemble model using a generalized additive model accounting for geographic differences to combine $PM_{2.5}$ estimates from three separate machine learning models; neural network, random forest, and gradient boosting. The three machine learning models complement each other; combining them using a generalized additive model provides an overall better model fit. Predictor variables in the three models included satellite data, meteorological variables, land-use variables, elevation, simulation outputs from chemical transport models, reanalysis data sets, and other data sources. Using the ensemble model, we predicted daily $PM_{2.5}$ from 2000 to 2016 at every 1 km $\times$ 1 km grid cell in the continental United States. Model training results for daily predictions from 2000 to 2016 indicated good model performance with a 10-fold cross-validated R^2 of 0.86. For annual estimates, the cross-validated R^2 was 0.89. The final model demonstrated good performance up to 100 $\mu g/m^3$. This work is completed, and a manuscript has been submitted to *Environment International*. A similar approach is being implemented for O_3.

In addition to $PM_{2.5}$ exposure estimates, we also estimated daily NO_2 concentrations from 2000 to 2016 in a similar ensemble model-based approach. Similarly, an ensemble model was fit using a generalized additive model to combine estimates from three machine-learning models; neural network, random forest, and gradient boosting, to obtain overall estimates of daily NO_2 concentrations. Predictor variables of the three machine-learning NO_2 models included NO_2 column concentrations from the satellite, land-use variables, meteorological variables, and other data sources. Using the ensemble model, we predicted daily NO_2 at 1 km $\times$ 1 km grid cells in the continental United States. The mean 10-fold cross-validated R^2 was 0.77, ranging from 0.67 to 0.79. The spatial R^2 (R^2 between monitored and predicted annual averages) was between 0.75 to 0.83, with a mean spatial R^2 of 0.82, indicating a good model performance at the annual level (unpublished results).

To summarize, compared with the existing model, our new ensemble $PM_{2.5}$ model performed noticeably better at the annual level. The existing model underestimated $PM_{2.5}$ concentrations at high concentrations. Our new ensemble model does improve model performance at high concentrations of $PM_{2.5}$. Furthermore, it is worth high-lighting that by

using the two new ensemble models we were able to estimate the uncertainty in the predictions (monthly standard deviation of the difference between daily monitored value and daily predicted value).

Applying New Causal Inference Methods to Same Data Used to Complete Prospective National Medicare Cohort Study

We are working on conducting causal inference analyses of the same data (for all the United States) and same study design as in Di and colleagues (2017c). Within this work, we are comparing various state-of-the-art approaches for causal inference to adjust for measured confounding. More specifically, we are comparing approaches using GPS models (IPTW, matching, doubly robust approaches) as well as nonparametric doubly robust approaches. In addition to adjusting for measured confounding, we are also exploring various approaches to adjust for unmeasured confounding by including time as a covariate in the GPS model. These planned analyses would more robustly explore methods to adjust for measured and unmeasured confounders by applying the innovative causal inference methods our group has been developing for Medicare data of the continental United States. In addition to our national Medicare cohort studies, we have begun applying new causal methods to Medicaid data from 2010 to 2011. Our first project using this cohort is to determine the short-term effects of $PM_{2.5}$ exposure on cardiovascular disease outcomes in this population. This is a new data set for our group to work with. We are using this first project to understand the challenges in analyzing such a complex data set, which includes varying amounts of data availability by state and differing eligibility criteria by state. Our first manuscript using this data is almost complete and will be presented at the next HEI meeting.

Discovering Heterogenous Groups under a Causal Inference Framework

We are developing causal inference methods for the de novo discovery of vulnerable subgroups and the estimation of air pollution effects within subgroups. Traditional air pollution health studies that seek to identify effect modifiers first select a priori a rather small set of these potential modifiers to test (either at the individual or area level) and then either (1) fit a regression model with one term for the exposure main effect and an additional interaction term between the exposure and each individual potential effect modifier, or (2) fit separate regression models stratified by the levels of a single potential modifier at a time. Recent studies have attempted to reduce the number of regression models by first identifying patterns across multiple modifiers, employing factor analytic techniques, and subsequently testing for potential modification by these composite factors (Achilleos et al. 2017; Kioumourtzoglou et al. 2016). Nonetheless, the current approaches have many limitations:

1. The potential effect modifiers to be tested are selected a priori, leading to the possibility of omitting key factors of vulnerability.
2. Inclusion of an interaction term in a regression model does not permit assessment of higher-order interactions (i.e., when a vulnerable subgroup is characterized by co-occurrence of multiple factors, which is likely among low income populations).
3. Although factor analytic and clustering approaches (Hastie et al. 2009; James et al. 2000) try to address this limitation, they are unsupervised approaches (i.e., one must first classify the population into subgroups and separately assess whether the risk varies by these previously identified subgroups). Thus, these approaches do not allow subgroup discovery based on the outcome and could potentially mask the characteristics of the truly vulnerable subgroups for the specific outcome of interest.
4. Previous techniques require solely continuous or categorical variables as inputs, which is not always the case when multiple modifiers are being evaluated.
5. The current epidemiological literature lacks state-of-the-art approaches to: (a) test for heterogeneity of air pollution effects across subgroups in a way that does not rely on strong parametric assumptions; (b) quantify the evidence of heterogeneity in a way that is not affected by residual or unmeasured confounding bias; (c) eliminate the multiple comparison problem; and (d) allow both continuous and categorical potential modifiers simultaneously as inputs. The new methods that we are developing are designed to overcome these limitations.

As part of our planned analyses, which are also summarized in a paper by Lee and colleagues (In press. *https://arxiv.org/pdf/ 1802.06710.pdf*), we developed a new approach for causal inference to discover *de novo subgroups of the population that experience causal effects of air pollution on mortality that are statistically significantly higher or lower than the population average.* The methods developed in this paper try to overcome the limitations of model misspecification, described as a limitation of Di and colleagues (2017c). More specifically, in this new approach for causal inference we split data into two subsamples, a training and a test sample: (a) in the training sample we use matching to eliminate measured confounding bias and consider a data-driven search for de novo discovery of subgroups that could have exposure effects that differ from the population mean; and then

(b) using the test sample, we quantify evidence of effect modification among the subgroups with nonparametric randomization-based tests. Because we first match and then test for effect modification using a nonparametric approach, we argue that this modeling design is more rigorous and less susceptible to false positive results for effect modification that could be driven by residual confounding bias. Via simulation studies and theoretical arguments, we demonstrate that since we discover the subgroups in the training set, hypothesis testing on the test set can focus on these subgroups only, thus substantially increasing the statistical power of the test. We have applied our method to the data from 1,612,414 Medicare beneficiaries in the New England region of the United States for the period 2000 to 2006. We find that low-income seniors 81–85 years and seniors over 85 years have statistically significant higher causal effects of exposure to $PM_{2.5}$ on 5-year mortality rates compared to the population mean. Scaling up this approach to the continental United States is challenging, but we intend to tackle this challenge as a primary goal of our planned analysis. The novel methods that we are developing, which will be generalizable and applicable to other analyses, will allow a granular joint identification of several potential effect modifiers.

Spatial Confounding and Analysis of Geographic Regions

In our completed work to date, we did not use geographic groups since we were interested in the overall effect of air pollution at the national level. However, it would be helpful for future analyses to study regional differences of the health effects of ambient air pollution. Such analyses could address residual confounding by geography. Further, it would be valuable to consider alternate geographic groupings to account for confounding (e.g., geographic areas that map more closely with patterning of disease rates in the United States, such as the so-called stroke belt in the southeastern United States).

IMPLICATIONS OF FINDINGS

Several critical factors position the scientific community at an unprecedented moment to accelerate scientific discovery and inform data-driven environmental and public health strategies. These factors include: incredible technological advances in how we collect data (from satellites, sensors, power plants, and electronic medical records); nearly unlimited computational power; and the development of new statistical methods that allow data to be analyzed in an unbiased, highly principled way to assess causality. However, in order to fully leverage these advancements to understand the impacts of long-term exposure to low levels of air pollution, we needed to address several existing gaps. Through our HEI-funded work, our team has fundamentally advanced the paradigm for scientific inquiry by: (a) developing a flexible R package called airpred (Sabath et al. 2018) that enables environmental health researchers to design and train spatio–temporal air pollution exposure models capable of predicting multiple pollutants, including $PM_{2.5}$; (b) developing new causal inference methods designed to account for exposure error and to improve ER estimation in order to account for differential confounding at different exposure levels. Together, these tools stand to significantly increase scientific rigor and advance evidence on the causal impacts of exposures to low levels of ambient pollution at a level or robustness not previously possible**.** The findings generated from our HEI-funded study have been, and will continue to be, impactful to the scientific community, policy makers, and the public.

ACKNOWLEDGMENTS

We would like to thank the Health Effects Institute for their support of our work, without which we would not have been able to advance the methods or evidence base at a scope and scale that is essential to understanding the health impacts of long-term exposure to ambient pollution.

We would also like to thank Harvard University for providing a stellar research computing environment. The computations in these articles were run on the Odyssey cluster, which is supported by the Harvard Faculty of Arts and Sciences' Division of Science, Research Computing Group, and on the Research Computing Environment, which is supported by the Institute for Quantitative Social Science in the Faculty of Arts and Sciences, both at Harvard University. Finally, we want to express our appreciation for Stacey Tobin and Amanda Spickard, who provided support in the preparation of this report.

REFERENCES

Achilleos S, Kioumourtzoglou MA, Wu CD, Schwartz JD, Koutrakis P, Papatheodorou SI. 2017. Acute effects of fine particulate matter constituents on mortality: A systematic review and meta-regression analysis. Environ Int 109:89–100; doi:10.1016/j.envint.2017.09.010

Andersen PK, Gill RD. 1982. Cox's regression model for counting processes: A large sample study. Ann Stat 10(4):1100–1120; doi:10.1214/aos/1176345976.

Carey IM, Atkinson RW, Kent AJ, van Staa T, Cook DG, Anderson HR. 2013. Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. Am J Respir Crit Care Med 187:1226–1233; PMID: 23590261; doi:10.1164/rccm.201210-1758OC

Cosier S. 2018. Clever use of public data could sidestep new rule. Science 360:473; PMID: 29724931; doi:10.1126/science.360.6388.473

Crouse DL, Peters PA, Hystad P, Brook JR, van Donkelaar A, Martin RV, et al. 2015. Ambient $PM_{2.5}$, O_3, and NO_2 exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). Environ Health Perspect 123:1180–1186; PMID: 26528712; doi:10.1289/ehp.1409276

Crouse DL, Peters PA, van Donkelaar A, Goldberg MS, Villeneuve PJ, Brion O, et al. 2012. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. Environ Health Perspect 120:708–714; PMID: 22313724; doi:10.1289/ehp.1104049

Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, et al. 2017a. Association of short-term exposure to air pollution with mortality in older adults. JAMA 318:2446–2456; PMID: 29279932; doi:10.1001/jama.2017.17923

Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J. 2016. Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. Environ Sci Technol 50:4712–4721; PMID: 27023334; doi:10.1021/acs.est.5b06121.

Di Q, Rowland S, Koutrakis P, Schwartz J. 2017b. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. J Air Waste Manag Assoc 67:39–52; PMID: 27332675; doi:10.1080/10962247.2016.1200159

Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. 2017c. Air pollution and mortality in the Medicare population. N Engl J Med 376:2513–2522; PMID: 28657878; doi:10.1056/NEJMoa1702747

Dockery DW, Pope CA, 3rd, Xu X, Spengler JD, Ware JH, Fay ME, et al. 1993. An association between air pollution and mortality in six U.S. cities. N Engl J Med 329:1753–1759; PMID: 8179653; doi:10.1056/NEJM199312093292401

Dominici F, Greenstone M, Sunstein CR. 2014. Science and regulation. Particulate matter matters. Science 344:257–259; PMID: 24744361; doi:10.1126/science.1247348

Hales S, Blakely T, Woodward A. 2012. Air pollution and mortality in New Zealand: Cohort study. J Epidemiol Community Health 66:468–473; PMID: 20966448; doi:10.1136/jech.2010.112490

Hastie T, Tibshirani R, Friedman J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York:Springer Science+Business Media.

James G, Witten D, Hastie T, Tibshirani R. 2000. An Introduction to Statistical Learning with Applications in R. New York:Springer Science+Business Media.

Kioumourtzoglou MA, Schwartz J, James P, Dominici F, Zanobetti A. 2016. $PM_{2.5}$ and mortality in 207 U.S. cities: Modification by temperature and city characteristics. Epidemiology 27(2):221–227;doi:10.1097/EDE.0000000000000422

Krewski D, Burnett RT, Goldberg MS, Hoover K, Siemiatycki J, Jerrett M, et al. 2000. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. A Special Report of the Institute's Particle Epidemiology Reanalysis Project. Cambridge, MA:Health Effects Institute.

Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, et al. 2009. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. Res Rep Health Eff Inst:5–114; discussion 115–136; PMID: 19627030. Available: *www.ncbi.nlm.nih.gov/pubmed/19627030*

Lee K, Small DS, Dominici F. In press. Discovering effect modification and randomization inference in air pollution studies. JASA Available: *https://arxiv.org/pdf/1802.06710.pdf.*

NSAPH (National Studies on Air Pollution and Health). Social-coding software platform. Available: *https://github.com/NSAPH/.*

NSAPH (National Studies on Air Pollution and Health). Open science framework. Available: *https://osf.io/2cg6v/.*

Ostro B, Hu J, Goldberg D, Reynolds P, Hertz A, Bernstein L, et al. 2015. Associations of mortality with long-term exposures to fine and ultrafine particles, species and sources: Results from the California Teachers Study Cohort. Environ Health Perspect 123:549–556; PMID: 25633926; doi:10.1289/ehp.1408565.

Pafka S. 2015. Benchmarking random forest implementations. Available: *http://datascience.la/benchmarking-random-forest-implementations/.*

R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. Vienna, Austria: The R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available: *http://www.R-project.org/.*

Sabath MB, Di Q, Schwartz J, Braun D, Dominici F, Choirat C. 2018. aipred: A flexible R package implementing methods for predicting air pollution. Presented at the 5th IEEE International Conference on Data Science and Advanced Analytics, 1–4 October 2018, Turin, Italy. Available: *arXiv:1805.11534v2 [stat.ML]* [accessed 3 July 2019].

Samet JM. 2011. The Clean Air Act and health — A clearer view from 2011. N Engl J Med 365:198–201; PMID: 21732828; doi:10.1056/NEJMp1103332

Shi L, Zanobetti A, Kloog I, Coull BA, Koutrakis P, Melly SJ, et al. 2016. Low-concentration $PM_{2.5}$ and mortality: Estimating acute and chronic effects in a population-based study. Environ Health Perspect 124:46–52; PMID: 26038801; doi:0.1289/ehp.1409111

Spiegelman D. 2016. Evaluating public health interventions: 4. The Nurses' Health Study and methods for eliminating bias attributable to measurement error and misclassification. Am J Public Health 106:1563–1566; PMID: 27509282; doi:10.2105/AJPH.2016.303377

Turner MC, Jerrett M, Pope CA, 3rd, Krewski D, Gapstur SM, Diver WR, et al. 2016. Long-term ozone exposure and mortality in a large prospective study. Am J Respir Crit Care Med 193:1134–1142; PMID: 26680605; doi:10.1164/rccm.201508-1633OC

Wang Y, Shi L, Lee M, Liu P, Di Q, Zanobetti A, et al. 2017. Long-term exposure to $PM_{2.5}$ and mortality among older adults in the southeastern US. Epidemiology 28:207–214; PMID: 28005571; doi:10.1097/EDE.0000000000000614

Wu X, Braun D, Kioumourtzoglou MA, Choirat C, Di Q, Dominici F. 2019. Causal inference in the context of an error prone exposure: Air pollution and mortality. Annals of Appl Stat 13(1):520–547.

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; PMID: 25399414; 10.1093/aje/kwu263

MATERIALS AVAILABLE ON THE HEI WEBSITE

Additional Materials contain the *New England Journal of Medicine* article by Di and colleagues (2017c) and the related supplementary appendix. It is available on the HEI website, *www.healtheffects.org/publications,* with permission from the publisher.

Additional Materials. Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. 2017. Air pollution and mortality in the Medicare population. N Engl J Med 376:2513–2522.

ABOUT THE AUTHORS

Francesca Dominici, PhD, is co-director of the Harvard Data Science Initiative and Clarence James Gamble Professor of Biostatistics, Population, and Data Science at the Harvard T.H. Chan School of Public Health. Her research areas include (1) linking massive and heterogeneous data sets to estimate health effects of environmental contaminants (air pollution, noise) and climate change; (2) developing statistical methods for the analysis of large and complex data to overcome the many sources of bias; (3) developing methods for causal inference to adjust for confounding bias and detect treatment effect heterogeneity; (4) applying newly developed methods to massive data sets to extract new knowledge regarding important questions in environmental health science, health-related impacts of climate change, and comparative effectiveness research; (5) leading the management and the analysis of several administrative databases, including Part A CMS files and SEER-Medicare, which are linked to air pollution and weather and socioeconomic data. She is an expert in Bayesian methods, longitudinal data analysis, confounding adjustment, causal inference, and Bayesian hierarchical models and data science and has extensive experience in the development of statistical methods and their applications to environmental epidemiology, implementation science and health policy, outcome research and patient safety, and comparative effectiveness research. She has authored more than 180 peer-reviewed publications and is a principal investigator, together with Dr. Xihong Lin, of an NCI P01 project titled "Statistical Informatics for Cancer Research." She is

also principal investigator of several U.S. EPA, National Institutes of Health, and HEI funded projects aimed at developing statistical methods and conducting nationwide epidemiological studies on the health effects of air pollution, and she oversees the management and analysis of several administrative databases, including Medicare and Medicaid claims data, which are linked to air pollution and weather and socioeconomic data. She is a permanent member of the National Institutes of Health study section of Biostatistical Methods and Study Designs, has served on a number of National Academies committees, and has recently been appointed as a member of the Radiation Effects Research Foundation Board of Councilors — a cooperative Japan–U.S. Research Organization focused on the study of health effects of radiation in the survivors of the atomic bombings of Hiroshima and Nagasaki. Dr. Dominici has received numerous recognitions, including the 2016 Janet L. Norwood Award for Outstanding Achievement in Statistical Sciences in 2016 and the 2015 Florence Nightingale David award, which is sponsored jointly by the Committee of Presidents of Statistical Societies and the Caucus for Women in Statistics. She was also awarded the 2013 Mathematics for Planet Earth Award Lecture, hosted by the Statistical and Applied Mathematical Sciences Institute; 2009 Diversity Recognition Award at Johns Hopkins University; the 2007 Myrto Lefkopoulou Distinguished Lectureship Award, Department of Biostatistics, Harvard T.H. Chan School of Public Health; 2007 Gertrude Cox Award, Washington, D.C., chapter of the American Statistical Association and RTI International; 2006 Mortimer Spiegelman Award, statistics section of the American Public Health Association; and Dean's Lecture at the Bloomberg School of Public Health in 2007.

Joel Schwartz, PhD, is professor of Environmental Epidemiology in the Department of Environmental Health at the Harvard T.H. Chan School of Public Health. He has more than 30 years of experience in the fields of epidemiology, exposure modeling, and biostatistics, including development of spatiotemporal statistical models that use satellite data to predict air pollutant concentrations. He is a worldwide expert and one of the most prolific scientists in the field of environmental epidemiology. He has extensive expertise in epidemiological methods and analyses looking at the health consequences of exposure to pollutants. His air pollution work has examined both acute and chronic effects of air pollution exposure. Recent research has established that exposure to fine combustion particles in the air at concentrations well below current standards are associated with a range of adverse health effects, from increased respiratory symptoms to increased hospital admissions to increased deaths. This work has led to a tightening of the U.S. air quality standards. He has done considerable work on health effects of O_3 exposure. He has several international collaborations underway in this area. His recent work has been focused on the cardiovascular effects of air pollution, and on factors which modify the response to air pollution. This work has suggested that people with diabetes are more susceptible, for example. He is also an expert in methods for causal inference and regression spline models, nonparametric smoothing, and generalized additive models. Dr. Schwartz also has extensive expertise in the use of cost–benefit analysis to make environmental decisions. He has developed benefit methodologies for assessing the benefits of lead control and has applied those methodologies to the decision to remove lead from gasoline, and recently, in collaboration with colleagues at the Centers for Disease Control, to a decision to revise their screening recommendations for children. He is also involved in cost–benefit analysis of air pollution control.

Qian Di, ScD, was a postdoctoral fellow in the Department of Environmental Health at Harvard T.H. Chan School of Public Health during the course of this study. He is currently assistant professor at the School of Medicine, Tsinghua University, Beijing, China. His work focuses on modeling, predicting air pollutants and temperature with high temporal and spatial resolutions, and working with health data to estimate health effects of air pollution and other environmental factors.

Danielle Braun, PhD, is a research scientist in the Department of Biostatistics at the Harvard T.H. Chan School of Public Health and Department of Computational Biology, working with Dr. Francesca Dominici, and also in the Dana-Farber Cancer Institute working with Dr. Giovanni Parmigiani. Her research focuses on the statistical development of methods in causal inference and risk prediction. She has worked extensively on measurement error, causal inference, comparative effectiveness research, risk prediction, genetic epidemiology, survival analysis, and frailty models. She has been mentoring undergraduate and graduate students for over four years and co-leads the BayesMendel lab along with Professor Parmigiani, leading many recent projects and working closely with PhD students on their theses.

Christine Choirat, PhD, was a senior research scientist in the Department of Biostatistics, Harvard T.H. Chan School of Public Health, Institute for Quantitative Social Science, Harvard University during the course of this study. She is currently chief health data scientist at Swiss Data Science Center, ETH Zürich and EPFL, Switzerland. A trained data scientist, Choirat has over 15 years of experience in computational statistics and in industry-standard software

development. Her research focuses on quantitative methods applied to economics, management, decision theory, and psychology, with a special emphasis on experimental approaches. She leads a big data acquisition, fusion, and archiving effort assembling a national database of ambient air quality monitoring, air pollution predictions, weather, population demographics, and Medicare/Medicaid hospitalization and mortality outcomes. She is a project lead of the popular statistical software, Zelig (*http://zeligproject.org/*), a suite of open-source R packages that facilitate using and interpreting a wide variety of statistical models through a user-friendly common interface. She has extensive experience in data repositories such as the Dataverse (*http://dataverse.org/*), and in fostering community engagement via the administration of User Groups (*https://groups.google.com/forum/#!forum/zelig-statistical-software*). She is co-author of genTB, an analysis tool for *Mycobacterium tuberculosis* genomic data.

Antonella Zanobetti, PhD, is a principal research scientist in the Department of Environment Health at the Harvard T.H. Chan School of Public Health. Her research focuses on the health consequences of exposure to air pollutants and climate change. Her interests are shaped by her background and training in both epidemiology and statistics and by her extensive experience developing and applying statistical methods for epidemiological investigations. Her work has contributed substantially to our understanding of air pollution and climate change-mediated health impact and is focused in three main areas: (1) examination of the impact of short- and long-term exposure to air pollution on mortality and morbidity and on cardiovascular disease in panel and longitudinal studies, (2) development of innovative statistical methodologies to examine emerging issues in environmental epidemiology, such as gene–environment interactions, epigenetics, and susceptibility and vulnerability, and (3) assessment of the health consequences of extreme temperatures and other weather parameters on mortality and morbidity. Dr. Zanobetti has recently been recognized in Thomson Reuters 2015 list of the most highly cited researchers, ranking in the top 1% of scientists cited in her field.

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Cutler D, Dominici F. 2018. A breath of bad air: Trump environmental agenda may lead to 80,000 extra deaths per decade. JAMA 319(22):2261–2262. PMID: 29896617, doi:10.1001/jama.2018.7351.

Zigler CM, Choirat C, Dominici F. 2018. Impact of National Ambient Air Quality Standards nonattainment designations on particulate pollution and health. Epidemiology 29(2):165–174. PMID: 29095246, doi:10.1097/EDE.0000000000000777.

Braun D, Gorfine M, Parmigiani G, Arvold N, Dominici F, Zigler C. 2017. Propensity scores with misclassified treatment assignment: A likelihood-based adjustment. Biostatistics 18(4):695–710. PMID: 28419189, doi:10.1093/biostatistics/kxx014.

Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, et al. 2017. Association of short-term exposure to air pollution with mortality in older adults. JAMA 318(24):2446–2456. PMID: 29279932, doi:10.1001/jama.2017.17923.

Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Dominici F, et al. 2017. Air pollution and mortality in the Medicare population. N Engl J Med 376:2513–2522. PMID: 28657878, doi:10.1056/NEJMoa1702747.

Dominici F, Zigler CM. 2017. Best practices for gauging evidence of causality in air pollution epidemiology. Am J Epidemiol 186(12):1303–1309. PMID: 29020141, doi:10.1093/aje/kwx307.

Makar M, Antonelli JL, Di Q, Cutler D, Schwartz J, Dominici F. 2017. Estimating the causal effect of low levels of fine particulate matter on hospitalization. Epidemiology 28(5):627–634. PMID: 28768298, doi:10.1097/EDE.0000000000000690.

COMMENTARY

HEI's Low-Exposure Epidemiology Studies Review Panel

Research Report 200, *Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1*, F. Dominici et al.

INTRODUCTION

This Commentary was prepared by the HEI Low-Exposure Epidemiology Studies Review Panel for the study "Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Pollution." This special Panel was convened to review HEI-funded studies on the health effects of exposure to low levels of ambient air pollution. The Commentary includes the scientific and regulatory background for the research, the Panel's evaluation of the Phase 1 report from the investigator team led by Dr. Francesca Dominici, and the Panel's conclusions. It is intended to aid the sponsors of HEI and the public by highlighting both the strengths and limitations of the study and by placing the Investigators' Report into scientific and regulatory perspective.

SCIENTIFIC AND REGULATORY BACKGROUND

The setting of ambient air quality standards — at levels considered adequate to protect public health — is a central component of programs designed to reduce air pollution and improve public health under the U.S. Clean Air Act (U.S. CAA*) and similar measures in Europe and around the world. Although the process for setting such standards varies, they all contain several common components:

- Identifying, reviewing, and synthesizing the scientific evidence on sources, exposures, and health effects of air pollution;
- Conducting risk and policy assessments to estimate what public health effects are likely to be seen at different levels of the standard;
- Identifying and setting standards based on scenarios considered in the risk analysis;
- Air quality monitoring to identify geographical areas that do not meet the standards; and,
- Implementing air quality control interventions to reduce ambient air concentrations to meet the standards.

SETTING NATIONAL AMBIENT AIR QUALITY STANDARDS UNDER THE U.S. CAA

The U.S. CAA requires that in setting the National Ambient Air Quality Standards (NAAQS), the U.S. Environmental Protection Agency (U.S. EPA) Administrator review all available science and set the NAAQS for all major ("criteria") pollutants (including ozone [O_3], particulate matter [PM], and nitrogen dioxide [NO_2]) at a level "requisite to protect the public health with an adequate margin of safety." In practice, since 2008 that review has had two principal steps:

1. Synthesis and evaluation of all new scientific evidence since the previous review in what is now called an *Integrated Science Assessment*. This document reviews the broad range of exposure, dosimetry, toxicology, mechanism, clinical research, and epidemiology evidence. It then — according to a predetermined set of criteria (U.S. EPA 2015) — draws on all lines of evidence to make a determination of whether the exposure is causal, likely to be causal, or suggestive for a series of health outcomes.
2. Assessment of the risks based on that science is then conducted in a *Risk and Policy Assessment*. This further analysis draws on the Integrated Science Assessment to identify the strongest evidence — most often from human clinical and epidemiological studies — of the lowest concentration levels at which health effects are observed, the likely implications of such levels for health across the population, and the degree to which the newest evidence suggests that there are effects observed below the then-current NAAQS for a particular pollutant.

Dr. Francesca Dominici's 4-year study, "Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Pollution," began in March 1, 2016. The Phase 1 draft Investigators' Report from Dominici and colleagues was received for review in October 2018. A revised report, received in February 2019, was accepted for publication in March 2019. During the review process, HEI's Low-Exposure Epidemiology Studies Review Panel and the investigators had the opportunity to exchange comments and to clarify issues in both the Investigators' Report and the Review Panel's Commentary. As the principal investigator of this study, Dr. Francesca Dominici, who is a member of the HEI Research Committee, was not involved in its selection for funding or in the oversight process.

* A list of abbreviations and other terms appears at the end of this volume.

> The Risk and Policy Assessment also examines the uncertainties around estimates of health impact and the shape of the concentration–response curve, especially at levels near and below the then-current NAAQS. Although a range of possible shapes of the concentration–response curves has been considered, including whether there is a threshold at a level below which effects are not likely, the U.S. EPA's conclusions in these reviews thus far have not found evidence of a threshold (although studies to date have not always had the power to detect one) (U.S. EPA 2004, 2013). Also, although the standard is set, according to statute, to protect public health with an adequate margin of safety, it has been generally understood that there are likely additional health effects below the NAAQS, although their presence and magnitude are more uncertain.

Both of these documents are subjected to extensive public comments and reviewed by the Clean Air Scientific Advisory Committee (CASAC), which was established under the U.S. CAA. CASAC is charged with peer-reviewing the documents — which includes providing guidance to the Administrator on the strength and uncertainties in the science and advising on alternative scenarios for retaining or changing the NAAQS.

EVOLUTION OF THE NAAQS

The reviews of the criteria pollutants have been ongoing for nearly 50 years, since the passage of the Clean Air Act Amendments of 1970. As the science has evolved, each subsequent review has examined the strength of the evidence for retaining or tightening the NAAQS. Although the process has frequently resulted in a decision to retain the then-current NAAQS, the NAAQS of both O_3 and fine PM (particulate matter $\leq$2.5 μm in aerodynamic diameter, or $PM_{2.5}$) have seen substantial revisions, especially over the last 20 years:

O_3 Starting in 1997, the NAAQS was converted from a 1-hour maximum standard to a standard averaged over 8 hours. In 1997, the NAAQS was set at 80 ppb; subsequently in 2008 it was lowered to 75 ppb, and then in 2015 to 70 ppb. Although there was epidemiological evidence of effects at or near these levels, the changes relied heavily on a series of carefully conducted human controlled-exposure studies.

$PM_{2.5}$ In 1997, based on dosimetric and biological information suggesting that fine particles less than or equal to 2.5 μg in diameter ($PM_{2.5}$) were a more appropriate indicator than PM_{10}, the U.S. EPA for the first time proposed and established a NAAQS for $PM_{2.5}$. It set the annual standard at 15 μg/m^3 in part as a result of the new long-term cohort evidence of association of $PM_{2.5}$ with adverse health effects (Dockery et al., 1993; Pope 1995) That was subsequently further reviewed in 2006 with no change and again in 2012, when the NAAQS, based on additional epidemiological evidence, was reduced to 12 μg/m^3 (U.S. EPA 2016).

IMPACT OF THE NAAQS

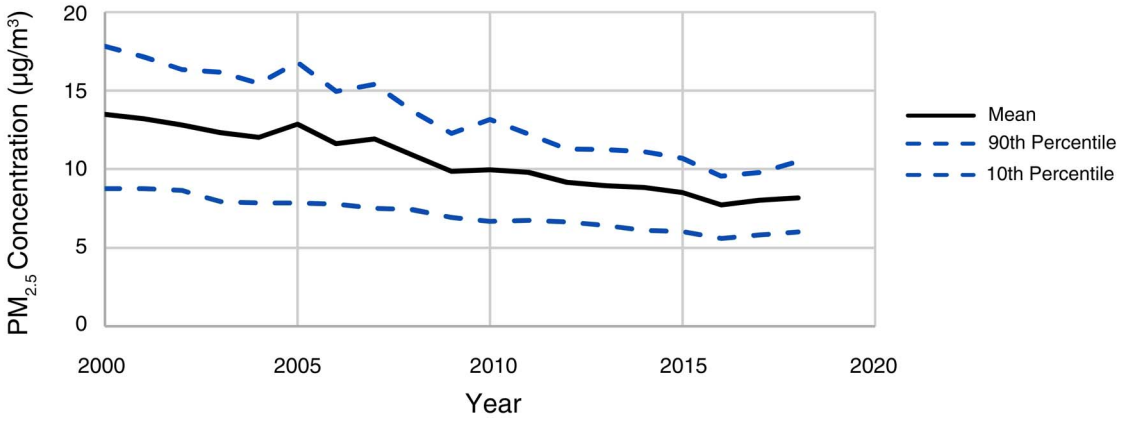
With the establishment of these standards, a host of national and regional regulatory actions began to reduce emissions from electric power plants, factories, motor vehicles, and other sources. As a result, there has been a steady and marked decline in ambient concentrations, so that much of the United States now attains the NAAQS (see, for example, the trend in $PM_{2.5}$ concentrations in the Commentary Figure.)

ADVENT OF RECENT STUDIES OBSERVING ASSOCIATIONS BELOW THE NAAQS

As the data on levels of $PM_{2.5}$ improved over the course of the first decade of this century, new studies began to emerge starting in 2012 (e.g., in Canada and New Zealand) suggesting that associations of $PM_{2.5}$ and mortality could be observed down to levels well below the NAAQS of 12 μg/m^3 (Crouse et al. 2012; Hales et al. 2012). These studies found robust associations, with some evidence of even steeper slopes of effect at the lowest levels, findings which, if replicated in other populations and by other investigators, could change the basis for future determinations of the levels at which to set the NAAQS and other air quality standards.

At the same time, they posed several questions, for example:

- Would the results be robust to the application of a range of alternative analytic models and their uncertainty?
- Could other important determinants of population health, such as age, socioeconomic position, health status, and access to medical care, as well as differences in air pollution sources and time–activity patterns, modify or confound the associations seen?
- Would the results change if risk estimates corrected for the effects of important potential confounding variables, such as smoking, in the absence of such data at the individual level?
- What might be the effects of co-occurring pollutants on health effect associations at low ambient concentrations?



Commentary Figure. Trends in $PM_{2.5}$ concentration from 2000 to 2018 (seasonally weighted annual average) as monitored by the U.S. EPA (data from U.S. *EPA; www.epa.gov/air-trends/particulate-matter-pm25-trends).*

As described in the Preface in this volume, the advent of these studies and the desire to address these important questions formed the basis for HEI's decision in 2014 to issue a Request for Applications (RFA 14-3), which sought and ultimately supported this study by Dr. Dominici and colleagues and two other studies that make up HEI's program to Assess Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution.

The Dominici research project encompasses a number of goals, which are addressed in nationwide exposure assessment and health impact studies, and causal modeling. Undergirding these studies is an effort to make the methods and data from this project available to the scientific community. The following evaluation is based on the initial results of the study described in the Phase 1 Investigators' Report.

SUMMARY OF THE STUDY

SPECIFIC AIMS

The full Dominici project, a four-year study funded by HEI, which began in 2016, has an expansive set of aims; however, for the purposes of this Phase 1 report, a subset of their overall aims is summarized as follows.

Aim 1: Exposure Prediction and Data Linkage Estimate long-term exposures to low levels of ambient $PM_{2.5}$ mass and the gaseous air pollutant O_3 by employing and extending hybrid prediction models that use satellite, land-use, emissions, ground-monitoring, and weather data, in conjunction with chemical transport models, at a high spatial resolution (1 km × 1 km) for the continental United States.

Aim 2: Causal Inference Methods for Exposure–Response Develop a new framework in Bayesian causal inference to estimate the concentration–response function that is robust to model misspecification for confounding and accounts for exposure error.

Aim 3: Evidence on Adverse Health Effects Estimate mortality associated with exposure to ambient air pollution for all U.S. Medicare enrollees between 2000 and 2012 (61 million adults, 65 years of age and older) and a representative subsample of Medicare participants with detailed personal information from the Medicare Current Beneficiary Survey (MCBS) (57,200 adults), using a cohort (long-term) and a case–crossover (short-term) design.

Aim 4: Tools for Data Access and Reproducibility Develop approaches and tools for data sharing, record linkage, and statistical software.

This commentary focuses in more detail on Aims 1 and 3, comprising initial results from the exposure and health effects research that have been published in peer-reviewed journals (Di et al. 2017a, 2017c; the latter can be found in Additional Materials on the HEI website). Aims 2 and 4 (causal modeling and more advanced statistical analyses and data access) are discussed briefly here as this research is still in its initial stages; the Panel does offer some

comments on this research, with suggestions for the conduct of these further analyses.

EXPOSURE AND HEALTH EFFECTS STUDIES

Data and Methods

The investigators amassed very large amounts of data from many different sources and used them for their analysis.

Ambient Air Pollution Concentrations Since the emphasis in this study was to study the entire older U.S. population — including people living in rural, low ambient air pollution concentration areas — the investigators developed air pollution concentration models for the 48 contiguous states, relying on research that they had completed before the current study began. They estimated ambient $PM_{2.5}$ concentrations for the period 2000 to 2012 using the following sources of data for their exposure model (for details, see Di et al. 2016):

1. Air monitoring data were obtained from the U.S. EPA Air Quality System (AQS), used in both model building and for cross-validation.
2. Aerosol optical depth (AOD) data were obtained from the moderate resolution imaging spectroradiometer (MODIS).
3. Surface reflectance data were also obtained from MODIS (MOD09A1).
4. Chemical transport model outputs were derived from the widely used GEOS-Chem model, which employs meteorological inputs and emission inventories to simulate atmospheric components. Total $PM_{2.5}$ was defined as the sum of nitrate, sulfate and ammonium ions and elemental carbon, organic carbon, sea salt aerosol, and dust aerosol. In addition to producing ground-level $PM_{2.5}$ estimates, the GEOS-Chem model is also useful for calibrating AOD because, being a three-dimensional model, it simulates vertical distribution of aerosols.
5. Meteorological data were obtained from the North American Regional Reanalysis project; the variables used included air temperature, accumulated total precipitation, downward shortwave radiation flux, accumulated total evaporation, planetary boundary layer height, low cloud area fraction, precipitable water for the entire atmosphere, pressure, specific humidity at 2 meters, visibility, wind speed, medium cloud area fraction, high cloud area fraction, and surface reflectance.
6. Aerosol index data were taken from the absorbing aerosol index measured by the ozone monitoring instrument (OMI), onboard the Aura satellite. These data are used to correct for the presence of other absorbing aerosols in the air (such as those from biomass burning and desert dust).
7. Land-use terms were obtained as previously described by Kloog and colleagues (2012). These terms represent emissions and can help inform small spatial scale variations; land-use data incorporate a variety of variables (such as population and road densities, emissions inventory, elevation, percentage urban, etc.).
8. In the regression models, the investigators also used regional and dummy variables to account for regional and temporal variability due to differences in meteorology and aerosol composition.

For estimating O_3 concentrations, the investigators used the same information for their models as listed for $PM_{2.5}$, supplemented by the following sources of data (see Di et al. 2017b):

1. Satellite-based O_3 measurements obtained from the OMI onboard the Aura satellite and used to calculate vertical distribution of O_3 levels.
2. Ozone vertical profile obtained through using an approach similar to that used for modeling $PM_{2.5}$. The GEOS-Chem model was used to estimate O_3 levels at different layers, and a scaling factor was used to calibrate satellite-based estimates to ground level O_3.
3. Ozone precursors (such as nitrogen oxides [NO_x], carbon monoxide, methane, and volatile organic compounds (VOCs), were estimated by the inclusion of AQS daily measurements of sulfur dioxide, NO_2, NO_x, and VOCs into the O_3 model, followed by the use of distance–decay functions from air quality monitors and other approaches.

With this large amount of data and using multiple approaches and input variables, the investigators developed a hybrid model to estimate daily $PM_{2.5}$ and O_3 levels at a 1 km × 1 km grid level. Complex atmospheric processes were addressed using a neural network that modeled nonlinearity and interactions. Spatial correlation was addressed using convolutional layers in the neural network, which aggregate nearby information and can simulate autocorrelation. The neural network was trained for the study period for the United States and tested against 10% left-out monitors. They then used the neural network to produce daily $PM_{2.5}$ levels (Di et al. 2016). Essentially the same approach was used to estimate and validate a model to predict daily O_3 concentrations during warm months (April 1 to September 30) (Di et al. 2017b).

Health Outcomes and Analyses Health data for this study were obtained from the Centers for Medicare and Medicaid Services (CMS), after applying through the Research Data Assistance Center (ResDAC) (*www.resdac.org*). The investigators obtained information on all Medicare beneficiaries for the years 2000 through 2012, which represents more than 96% of the U.S. population 65 years of age or older. This is an open cohort where individuals enter when they enroll in Medicare at or after age 65 and stay until death. Individuals with an unverified date of death were excluded. For each beneficiary, the following data were extracted: the date of death (if applicable), age at year of Medicare entry, calendar year of entry, sex, race, ethnicity, ZIP code of residence, and Medicaid eligibility (a proxy for low socioeconomic status (SES); note that these individuals were eligible for both Medicare and Medicaid). Thus, all deaths among Medicare recipients during 2000 to 2012 were captured. In all, the cohort had about 61 million persons, with 460 million person-years of follow-up and 23 million deaths.

Medicare data contain little information about individual-level covariates. Therefore, the investigators also obtained data from the MCBS, which is an annual phone survey of a nationally representative sample of Medicare beneficiaries and contains information on more than 150 potential individual confounders, including data on individual risk factors (e.g., smoking, body mass index [BMI], and income). Information on a sample of more than 57,000 enrollees was obtained for the period 2000 through 2012. Dominici and colleagues also analyzed data for a cohort of ~32,000 beneficiaries from the MCBS-Medicare database, which links data from MCBS interviews with Medicare claims data, and also contains information on confounders (see Di et al 2017c, Supplementary Appendix, Section 5, found in Additional Materials on the HEI website; and Makar et al. 2017). The Commentary Table is a summary of the potential confounders that were used during this study.

The investigators used both cohort and case–crossover designs to analyze the association between exposure to $PM_{2.5}$ and O_3 and all-cause mortality in the Medicare cohort from 2000 to 2012. For the cohort study, they performed survival analyses using the Andersen-Gill (AG) method (Andersen and Gill 1982), a variant of the traditional Cox proportional hazards model that incorporates spatiotemporal features by allowing for covariates to vary from year to year. They estimated hazard ratios associated with a 10-$\mu g/m^3$ increase in $PM_{2.5}$ and a 10-ppb increase in O_3 exposure using this model in two-pollutant analyses.

The investigators developed concentration–response curves for air pollution levels and mortality by fitting a log-linear model with thin-plate splines of both pollutants while controlling for all individual and ecological variables that they had used in their main analyses (details in Di et al. 2017c, Supplementary Appendix, Section 7, found in Additional Materials). In view of the objective of this research, they explored the health effects at lower ambient concentrations by conducting separate analyses that included only person-years with $PM_{2.5}$ exposures lower than 12 $\mu g/m^3$ and O_3 exposures lower than 50 ppb.

To assess if any subgroups within the Medicare cohort were at higher or lower risk of mortality associated with air pollution, the investigators fitted the same Cox model as above for certain subgroups (e.g., male vs. female, white vs. black, and Medicaid eligible vs. Medicaid ineligible). To explore the robustness of the results, they performed sensitivity analyses and compared any changes in risk estimates with differences in confounder adjustment and estimation approaches. Finally, since Medicare data do not include information on many important individual-level covariates, the investigators utilized data from the MCBS. Using individual-level data (such as smoking status, BMI, and income) and data on many other covariates from the MCBS, they examined how the lack of adjustment for these risk factors could have affected the risk estimated for the Medicare cohort (Di et al. 2017c, Supplementary Appendix, Section 5, found in Additional Materials).

For the case–crossover study, the case day was defined as the date of death; the daily exposure to air pollution for the case day was defined as the mean of the ambient concentration on that day and the day before (i.e., mean of lag 0-day and lag −1-day). For each person, they compared daily air pollution concentration on the case day versus daily air pollution exposure on control days, which were chosen (1) on the same day of the week as the case day, to control for potential confounding effect by day of week; (2) before and after the case day to control for time trend; and (3) in the same month as the case day to control for seasonal and subseasonal patterns. They fitted a conditional logistic regression to all pairs of case and matched control days, thus estimating the relative risk of all-cause mortality associated with short-term $PM_{2.5}$ and O_3 exposure (Di et al. 2017a).

The investigators controlled for potential residual confounding by weather-related factors by using natural splines of air and dew point temperatures with 3 degrees of freedom. For subgroup analyses, the investigators used information on sex, race, or ethnicity (white, nonwhite, and others), age categories (≤ 69, 70–74, 75–84, and ≥ 85 years), eligibility for Medicaid, and population density at residence (in quartiles). Subgroup-specific estimates of relative risk and absolute risk difference were obtained by fitting separate conditional logistical regression models to

Commentary Table. Characteristics of Potential Confounders and Variables[a]

Potential Confounder / Model Covariate	Variable Level	Variable Type	Data Source[b]
Age Age at entry	Individual	Categorical (5-year)	Medicare
Race White Black Asian Hispanic) Native American	Individual	Binary and continuous (% of population)	Medicare and U.S. Census, ACS
Sex Sex	Individual	Binary	Medicare
Smoking Ever smoker (%)	Ecological (county to ZIP code)	Proportion	BRFSS (2000–2012)
Obesity BMI	Ecological (county to ZIP code)	Continuous	BRFSS (2000–2012)
Diet Not included	n/a	n/a	n/a
Exercise Not included	n/a	n/a	n/a
Socioeconomic status — individual level Medicaid eligibility	Individual	Binary	Medicaid Statistical Information System
Socioeconomic status — community level Median household income Median value of housing % owner occupied % below poverty level (age >65) % below high school education (age >65) Population density	Ecological (ZIP code)	Continuous	U.S. Census, ACS
Access to health care % with LDL-C % with HgbA1c test % with ≥ 1 visit	Ecological (ZIP code)	Continuous	Dartmouth Atlas of Health Care
Meteorological Temperature Relative humidity	Area (32 km × 32 km)	Continuous	North American Regional Reanalysis data
Regional dummy variable 10 geographical regions with similar $PM_{2.5}$ chemical profile	Regional	Categorical	GEOS-Chem 3D global chemical transport model

[a] Based on information in Di et al. 2017c, Supplementary Materials (available in Additional Materials on the HEI website).

[b] American Community Survey (ACS); Behavioral Risk Factor Surveillance System (BRFSS); n/a = not applicable.

the data for each subgroup. To test for statistically significant differences in estimated relative risk and the absolute risk difference between categories within each subgroup (e.g., male vs. female), they used a two-sample test, based on the point estimate and standard error. They explored the health effects at lower levels of exposure by performing subanalyses with cases restricted to those occurring on days with daily air pollution concentrations below 25 μg/m^3 for $PM_{2.5}$ and 60 ppb for O_3.

Key Findings Reported by the Investigators

Exposure Assessment Dominici and colleagues reported overall good performance for the $PM_{2.5}$ prediction model, with R^2 of 0.84 (range 0.78 to 0.88) (Di et al. 2016). During the course of the study, annual $PM_{2.5}$ concentrations ranged from 6.2 to 15.6 μg/m^3 (5th and 95th percentiles, respectively). The average annual $PM_{2.5}$ concentration was 11.0 μg/m^3 during the study period, 2000–2012. The model performed better in the eastern and central United States and less well in the western United States (Di et al. 2016, Supplementary Appendix Table S4). The highest $PM_{2.5}$ concentrations were predicted to be in the eastern and southeastern United States and in parts of California. The R^2 values were lower after 2010, apparently as $PM_{2.5}$ ambient levels decreased in the eastern United States (R^2 in 2000 and 2001 of 0.86 and 0.84 vs. 0.81 and 0.74 in 2011 and 2012). In addition, the model performed better during the summer — when $PM_{2.5}$ levels often tend to peak — followed by autumn, spring, and winter (mean R^2 values of 0.88, 0.84, 0.84, and 0.80, respectively) (Di et al. 2016, supplementary materials).

The O_3 prediction model performed similarly well, with an overall R^2 of 0.80 (Di et al. 2017b).* The average of 8-hour, daily, warm-season O_3 concentrations across the country during the study period ranged from 36 to 56 ppb (5th and 95th percentiles, respectively), with an average of 46.3 ppb during the study period. The investigators found a west–east gradient in the O_3 level, with the model performance being the best in the middle Atlantic, south Atlantic, east north Central, west south Central, and the Pacific States regions. Model performance was not affected by the year, so no year-to-year trend in model fit was observed. Seasonal trends in model performance were also apparent, with the R^2 being highest in the autumn, followed by summer, spring, and winter (R^2 values of 0.75, 0.71, 0.68, and 0.67, respectively). O_3 concentrations were the highest in the Mountain region and in California and were lower in the eastern states. Annual $PM_{2.5}$ and warm-season O_3 concentrations were only weakly correlated, with a Pearson correlation coefficient of 0.24.

The Cohort Study The 2000–2012 cohort of Medicare beneficiaries, with about 61 million enrollees and 23 million deaths, provided a very large population to study association with the long-term exposure to ambient air pollution, including at concentrations below the current NAAQS for both $PM_{2.5}$ and O_3. In two-pollutant analyses, Dominici and colleagues report a 7.3% (95% confidence interval [CI], 7.1% to 7.5%) higher risk of all-cause mortality for each 10-μg/m^3 increase in annual average $PM_{2.5}$ concentrations and a 1.1% (CI, 1.0% to 1.2%) higher risk of mortality for each 10-ppb increase in annual average O_3 concentration in the warm season (Di et al. 2017c). At low concentrations — less than 12 μg/m^3 for $PM_{2.5}$ and less than 50 ppb for O_3 — the risk was 13.6% (CI, 13.1% to 14.1%) for $PM_{2.5}$ and 1.0% (CI, 0.9% to 1.1%) for O_3. Thin-plate-spline regression analysis for concentration–response relationship in two-pollutant models produced almost linear curves, with no suggestion of a threshold down to 5 μg/m^3 of PM and 30 ppb of O_3 (see Figure 7 in the Investigators' Report and Supplementary Appendix, Section 5 of Di et al. 2017c, found in Additional Materials).

In subgroup analyses for $PM_{2.5}$, the investigators found larger estimates of effect among males and among Hispanics, Asians, and particularly African Americans compared with whites. Individuals with low SES, as indicated by eligibility for Medicaid, appear to have a slightly higher risk per unit of air pollution (Di et al. 2017c, Supplementary Appendix, Table S3, found in Additional Materials). For long-term O_3 exposure, the subgroup analysis showed that the effect estimates were higher for Medicaid-eligible enrollees and slightly higher for whites, but these analyses also produced hazard ratios of less than 1 for certain subgroups, including Hispanics and Asians, and particularly for Native Americans, than the overall population.

The Case–Crossover Study The case–crossover analyses comprised more than 22 million deaths (case days) and more than 76 million control days among Medicare enrollees between 2000 and 2012, again a very large population. For short-term exposures, the investigators observed a 1.05% (95% CI, 0.95% to 1.15%) greater risk of mortality in two-pollutant models for a 10-μg/m^3 increase in $PM_{2.5}$ concentrations, and 0.51% (CI, 0.41% to 0.61%) for a 10-ppb increase in average 8-hour warm-season O_3 concentration (pollutant levels were averaged over the current and previous day) (Di et al. 2017a). At low concentrations (<25 μg/m^3 of $PM_{2.5}$ and <60 ppb of O_3), the associations remained elevated for both $PM_{2.5}$ and O_3, with relative risk

*Note that this publication reports the R^2 as 0.76 [range 0.74 to 0.80]; presumably this is because the authors improved the model after publication of the earlier article (Di et al. 2017b).

increases (RRI) of 1.61% (95% CI, 1.48% to 1.74%) and 0.58% (CI, 0.46% to 0.70%), respectively. In exposure–response curves, the relative risk increase rises sharply for both pollutants at a relatively low concentrations and then levels out at higher concentrations (see Figure 8 in the Investigators' Report).

In subgroup analyses for the case–crossover study, significant effect modifications were reported for several variables. For $PM_{2.5}$, the investigators observed higher mortality risk for females and individuals who were older (age >70 years), black, or eligible for Medicaid (i.e., lower SES) (Di et al. 2017a, Figure 3). For O_3, there was much less contrast between groups, except for age where the older group had a significantly higher risk of mortality (0.69 for ≤69 years vs. 1.83 for ≥85 years) (Di et al. 2017a, Figure 4).

REVIEW PANEL EVALUATION

This report by Dominici and colleagues summarizes an impressive amount of work completed in the first part of this project. There are several particularly strong aspects of this work: The investigators amassed an extremely large cohort by compiling a very large amount of data on health and related factors across the continental United States from national databases (Medicare and others). They also estimated U.S.-wide air pollution concentrations at high spatial resolution (with 1 km × 1 km grids) and temporal resolution (enabling daily averages). Finally, they developed and applied state-of-the-art statistical techniques to the assessment of health effects of low levels of air pollution.

The Panel's evaluation of this report was made challenging by the nature of the report submitted; the Phase 1 study report was largely compiled from the initial published reports, as well as from some as-yet-unpublished methodological work. The Panel has therefore expanded the focus of this review to include — in addition to the Investigators' Report — some of this recently published work (in particular, Di et al. 2017a and 2017c; the latter can be found in Additional Materials on the HEI website). Di and colleagues have provided many details in the supplemental materials of the two publications. Additionally, the Panel communicated with the investigators during the course of the review. In response to comments from the Panel, the investigators added an additional discussion to the Investigators' Report of limitations and plans for future work.

As stated earlier, the Phase 1 report represents a snapshot of the ambitious work undertaken by the investigators. Much work, including further development of causal methods that would properly allow for the complexities in the design of the studies and nature of the data is currently ongoing. As a whole, this work is likely to represent an important contribution to the literature on the health impacts of air pollution on older adults in the United States. The current report represents a high-quality and thorough investigation of some of the most challenging problems in environmental health.

EXPOSURE ASSESSMENT

The use of large, diverse, and existing data sets to generate estimates of $PM_{2.5}$ and O_3 concentrations on a 1 km × 1 km national grid for the entire continental United States (~8 million km^2) is impressive, both in terms of the vast amount and variety of data assembled and the tremendous computational requirements for the analysis (Di et al. 2016, 2017b). The methods developed should prove valuable to researchers studying air pollution and health, especially because the investigators have made efforts to make their modeling approach publicly available for others to use.

Using a hybrid model, Dominici and colleagues estimated $PM_{2.5}$ and O_3 concentrations in areas where monitors are sparse, allowing estimates for a larger number of ZIP codes, and thus individuals, to be included within the analyses. However, as with any exposure assessment, it is critical to consider the potential for prediction errors, particularly those that may be systematic, and the implications for the interpretation of the associated epidemiological results. Specific strengths and weaknesses of the exposure assessment are discussed below.

First, Dominici and colleagues used U.S. EPA ground-monitoring data to cross-validate their exposure models. Regional and monthly dummy variables were used in the model in an attempt to account for regional and daily variations related to differences in meteorology and aerosol composition (Di et al. 2016). However, both geographical and temporal variability in the errors of the concentration estimates remained in the final estimates for both $PM_{2.5}$ and O_3, as discussed earlier. The source(s) and impact of such variability are not understood and deserve attention.

Second, because U.S. EPA monitors are located for the purpose of compliance with NAAQS, they are generally placed in the more populated, urban areas where air pollution levels are higher. Consequently, the rural areas — where population density is lower and lower pollutant concentrations are found — are not as intensively monitored, and the model may be more prone to larger error in such areas. Further, rural ZIP codes generally cover much larger areas than urban ZIP codes. The potential impact of this on exposure estimates can be seen, for example, in the lower R^2 values for $PM_{2.5}$ estimates for the Mountain region (see Di et al. 2016, Figure 1). Although only about 25% of the U.S. population lives 20 km or farther away

from the nearest monitoring station — primarily in rural areas — these are the residents of potentially greatest interest, in the context of this study, because of their lower exposures to pollutants; therefore, the nature, size, and potential impact of these errors are important to understand.

Third, based on the relationship between the model predictions and observed $PM_{2.5}$ and O_3 levels (see Di et al. 2016, Figure 5, and Di et al. 2017b, Figure 6), it appears that the model may systematically underpredict concentrations (i.e., produce predictions below the 1:1 line). The impact of such underprediction may be important and should be explored in future research. (Both curves show much greater uncertainty at high pollutant concentrations, but few people live in such high-concentration areas.)

Finally, though the Panel recognizes that the investigators were building a very large, national-scale model with a resolution of 1 km × 1 km, the model does not capture fine scale variability in ambient concentrations. Thus, the model at this scale does not capture local, high gradients in concentrations, such as those along roadways or near major point sources. The exposure estimation for those living in the vicinity of such areas is probably underestimated (for $PM_{2.5}$) or overestimated (for O_3, because of local area scavenging), though typically $PM_{2.5}$ and O_3 levels tend to be more uniform at urban and regional scales than pollutants such as NO_2, which exhibit higher spatial variation.

Using input from disparate sources to develop a model at the national scale, with a 1 km × 1 km resolution, is a major accomplishment, though the model has its limitations. The Panel has noted that the investigators are taking steps to improve their models — by using three different machine-learning models that complement one another — and extending their models to the year 2016. In addition to updating the $PM_{2.5}$ and O_3 models, they are also modeling NO_2 (see the "Next Steps" section in the Investigators' Report). The application of the improved and additional models for epidemiological analysis should prove useful and may shed greater light on the exposure–response relationships described in these two studies.

HEALTH EFFECTS: COHORT STUDY*

Using the massive database of all Medicare recipients during 2000 to 2012, and combining it with the equally large exposure predictions, Dominici and colleagues have performed a study with unsurpassed power to investigate the association between all-cause mortality and long-term exposure to a range of $PM_{2.5}$ and O_3 levels. That they observed an association between annual average concentrations and mortality at higher concentrations was not the new finding of this work, but the findings at low levels, particularly at levels below the current NAAQSs, are novel and potentially important.

The greatest challenge to the internal validity of this study, as for all observational studies, is the potential for confounding, which can bias the results. To address such concerns, the investigators performed numerous analyses with some 20 covariates (Commentary Table) (for details, see Di et al. 2017c, Supplementary Appendix, available in Additional Materials on the HEI website). They also utilized findings from a smaller Medicare cohort that had a much richer set of potential confounding variables to assess the likely impact of having only a limited number of covariates in the main cohort analysis. To allow for the effects of time-dependent covariates that are known to vary from year to year, the investigators utilized a variant of the classic Cox proportional hazards model — the AG formulation (Andersen and Gill 1982).

However, this is a complex study. Health and personal characteristics are available for individuals, but ambient air pollutant exposure is estimated at the ZIP code level (averaged from the 1 km × 1 km spatial scale of the prediction model). Additionally, the ZIP code scale is the smallest spatial unit at which individual residential and other covariate information is available. These factors, coupled with confounders that can act at the level of the individual, the community, or the regional environment, result in a complex hybrid model. These issues pose important challenges for the next phase of the work planned by the investigators, and the causal inference methods under development will need to focus on these challenges. Based on the current results, the Panel offers the following comments.

Temporal Confounding

Although the investigators have used the AG formulation of the Cox proportional hazards model to better represent time-dependent variables, the Panel's biggest concern relates to the problem of the potential for temporal confounding, with both the overall nonaccidental mortality and the $PM_{2.5}$ levels declining steadily over the period of the study, 2000 to 2012. Since this is an open cohort (new individuals enter the cohort as they enroll for Medicare), age — which is controlled in the analyses — is not necessarily strongly correlated with calendar time. As a result, confounding can occur due to the contributions of both age and calendar time. In this study, however, there was no adjustment for calendar time, and age was included in the models using five-year categories. Although the Panel understands that there are computational challenges to

* A copy of the article by Di and colleagues (2017c), along with its Supplementary Appendix, is available in Additional Materials on the HEI website, with permission of the publisher.

including a finer resolution for age, the supplementary materials accompanying the article by Di and colleagues (2017c, Supplementary Appendix, found in Additional Materials) show that the hazard ratio drops from 1.07 when a five-year age category is used to 1.05 when it is replaced with a three-year age category. This suggests that this question is unresolved and deserves more attention. Similarly, the Panel acknowledges that disentangling secular trends from any possible causal effect of $PM_{2.5}$ on mortality can be challenging and that including year in the models may over-adjust for exposure by removing true variability over time. Regardless, the inability to adequately account for potential bias due to temporal trends introduces a large element of uncertainty in interpreting the study's findings to date.

In summary, the Panel believes that, without accounting for confounding by time, the findings of the long-term exposure study should be viewed with caution. The Panel is glad to note that the investigators acknowledge these limitations and looks forward to the development of appropriate causal inference techniques and their application to the Medicare data set.

Potential for Residual Confounding

Dominici and colleagues have performed various analyses to explore the possible sources of residual confounding; however, as discussed below, the Panel identified several areas with a potential for residual confounding in the cohort study that need resolution.

Subgroup Differences Some results from the subgroup analyses are puzzling, as acknowledged by the investigators' team: for example, the dramatically higher effect of $PM_{2.5}$ in African Americans and the negative (protective) effects of O_3 for Native Americans, Hispanics, and Asians. It is possible that these observations reflect true intergroup differences; alternatively, it may be more likely that the subgroup designation serves as a surrogate for other risk factors not fully considered, resulting in residual confounding. Model misspecification is another possibility.

Spatial Differences Another issue here is the different scales at which the exposure and health models operate. The Panel has concerns about the impact of the likely exposure misclassification and confounding related to the spatial differences between aggregated summaries of exposures (1 km × 1 km) and residential locations (at the ZIP code level). The Panel appreciates that the health and covariate data are available only as aggregated ZIP-code-level values and looks forward to the results of the planned analyses in the Final Phase 2 report, in which the investigators plan to explore exposure measurement error in the health analyses using a causal-inference framework. The Panel is also aware that the exposure measurement error correction methodology for spatially varying pollutants in multipollutant research is in its infancy (e.g., Bergen et al. 2016; Szpiro and Paciorek 2013), and even more so in the causal inference framework — as duly acknowledged by the investigators — so it is not surprising that Dominici and colleagues did not yet address this in their extensive work.

Smoking, Diet, and Exercise Data on individual health-related behaviors, which are well known for affecting survival time, were available only at the ZIP code level. Some of the information — for example, binary variables for smoking behavior — does not capture the full extent of the variability in the behaviors. The Panel understands the complexity of these factors and the difficulty in finding data on a national scale to include in the model. However, some of these behaviors are known to vary regionally, and it is conceivable that one or the other is geographically correlated with $PM_{2.5}$ or O_3. For example, residents of the southeast have some of the highest $PM_{2.5}$ exposure levels and also have the highest rates of obesity in the United States (Centers for Disease Control and Prevention 2019).

Socioeconomic Status (SES) The investigators appropriately consider a variety of measures of SES at the individual and community level; these measures represent a number of factors that might increase mortality risk. They include baseline health status, diet, exercise, psychosocial stressors, risk of violent crime, risk of exposure to chemical and microbial contaminants, and access to medical care. The only measure of individual-level SES available for the entire cohort is Medicaid eligibility status, which produced a fairly small difference in hazard ratios (eligible 1.080 vs. noneligible 1.075) (Di et al. 2017c, Supplementary Appendix, Table S3, found in Additional Materials). To the extent that Medicaid eligibility is an imperfect measure of the relevant aspects of SES, additional sources for residual confounding may be present.

The issues with individual-level SES notwithstanding, neighborhood SES factors — not individual SES — have been reported to be the more important confounders affecting air-pollution-associated mortality (Hajat et al. 2013; Makar et al. 2017). The investigators used four different and reasonable measures of community SES: median household income, median housing price, percentage below poverty level, percentage of homes owner-occupied, and percentage below high school education. They report that none of these had a significant correlation with the observed outcomes (Di et al. 2017c, Supplementary Appendix, found in Additional Materials). The adjustment for neighborhood

SES partly addressed concerns about the limitations of accounting only for individual-level SES, so the inclusion of these additional SES-related factors in the analyses is a strength of this study.

Cohort vs. Case–Crossover Analysis The Panel was not persuaded by the claim made in the "Limitations" section of the Investigators' Report that estimating effects in both the case–crossover and cohort analyses provides some assurance against confounding. At best, this provides evidence that PM does affect mortality. However, the nature of the confounders and the effects being estimated (Eftim and Dominici 2005; Künzli et al. 2001; Rabl 2003) are so different that consistency of findings across the two designs provides essentially no assurance against confounding.

Precision of Effect Estimates

Another issue to consider is related to one of the major strengths of the study: the extremely large number of observations. Statistical methods have been developed in light of the limitation that an entire population is generally not available for study, so one must study a sample of the population. Statistical methods related to the estimation of different parameters (e.g., bias) and related inferences (e.g., CIs and *P* values) are based on the premise that study participants are sampled from a larger existing or theoretical population. The Dominici study represents a growing trend in the new "Big Data" era in that the *entire* Medicare population of more than 60 million individuals has been studied. Though this enormous sample gave the study unprecedented power to investigate effects, it also raises questions about interpretation of the very narrow CIs and other comparative statistics reported for the cohort. In this situation, bias and model misspecification are likely to be more critical concerns than sampling variability. Because the impact of bias and model misspecification is not reflected in standard uncertainty measures, one should be cautious about overinterpreting the narrow CIs, as the interval width is driven by the very large sample size (see Meng 2018), and the Panel's comments and concerns about the potential impacts of bias and of unmeasured confounding should be viewed in this broader context.

Other Pollutants

Dominici and colleagues have looked at mortality associations with both $PM_{2.5}$ and O_3; this is another strength of this study. However, other pollutants may also confound the associations between $PM_{2.5}$ and O_3 and mortality. The Panel looks forward to the results of ongoing work to strengthen the current exposure models (e.g., using data from the IMPROVE network) and to the inclusion of a NO_2 model and possibly PM composition.

HEALTH EFFECTS: CASE–CROSSOVER STUDY*

Long-term studies are typically considered more important for risk and burden assessments as well as policy making, though short-term studies have played an important role as well in the development of air pollution epidemiology science and its applications to policy. The second epidemiology study in this report uses a case–crossover design — a variant of the time-series design — with the Medicare population to evaluate short-term effects of air pollution exposure. One advantage that this design has over the study of long-term health effects is that it is based on variation in exposure and mortality over short periods of time (days, rather than years). Therefore, only confounding factors that vary over short periods of time, such as weather, are of potential concern, rather than the much larger array of potential confounders that either do not vary with time or have long-term trends. On the other hand, by design time-series analyses address only the immediate impact of air pollution on mortality rather than the role of pollutants in the development of chronic morbidity and subsequent mortality. The two designs are both valuable analyses but address different sets of covariates and different questions.

Dominici and colleagues report an RRI of 1.05% (95% CI, 0.95%–1.15%) and 0.51% (0.41%–0.61%) in daily mortality rate, respectively, for each 10-μg/m^3 increase in $PM_{2.5}$ and 10-ppb increase in O_3 (Di et al. 2017a). The concentration–response analysis for $PM_{2.5}$ and O_3 suggests a nonlinear relationship, with a steeper slope at low concentrations and flattening at higher concentrations (see Figure 8, Investigators' Report). The investigators have provided the effect estimates for concentrations below 25 μg/m^3 for $PM_{2.5}$ and 60 ppb for O_3, which are the concentrations of interest for this study and below which the curves are linear.

In addition to the main findings, the authors have investigated effect modification for a range of variables. For example, they report that the mortality effect of short-term exposure to $PM_{2.5}$ is greater in women than in men (RRI of 1.20 vs. 0.86; Di et al. 2017a, Figure 3), in contrast to the finding in their cohort study. There is again a clear age effect, particularly for O_3 exposure, with older individuals having a significantly higher RRI. The effects in other subgroup analyses were generally not significant, except for Medicaid eligibility. An important group of time-varying covariates not fully included in these models is copollutants, such as NO_2.

* The paper by Di and colleagues (2017a) may be viewed at *https://jamanetwork.com/journals/jama/fullarticle/2667069* (courtesy of JAMA).

SHARING OF MODELS AND DATA

From the inception of this project, the Panel was glad to note that the investigators planned to make available their methods, models, and data with other investigators. To facilitate this, they have posted their data, workflows, and analyses to a secure high-performance computing cluster with the objective of developing an open science research data platform (*https://osf.io/2cg6v/*). Additionally, the codes and software tools are available from the location *https://github.com/NSAPH/airpred*. The investigators' efforts in this area — to make both models and data available — will continue.

Model With an interest in making their model widely available, the investigators developed a flexible R package so that interested environmental health scientists may design and train spatiotemporal models that can predict air pollutants, including $PM_{2.5}$ (Sabath et al. 2018). This is accomplished via neural network tools to produce exposure predictions with high spatial (1 km × 1 km grids) and temporal (enabling daily averages) resolution. The adoption of the R platform — as opposed to the less user-friendly MATLAB platform used by Di and colleagues (2016, 2017b) in their work and on which the airpred package is based — is a major strength since it is likely to promote wider use of the modeling tools by other environmental health researchers. The use of an open source big data platform (H2O) for better computational efficiency and hence scalability is also another major strength. The R package airpred has the flexibility to allow specification of "different types of neural networks, with different parameters, or even to perform ensemble modelling."

Data In their research, Dominici and colleagues have made use of a great deal of data generated by public sources, including the National Aeronautics and Space Administration, the U.S. EPA, and the CMS; data from most of these sources are in the public domain and readily available to anyone. The one exception is the Medicare data, which the investigators are prohibited from sharing under terms of access of the data from CMS. However, these data are available from ResDAC; following an application, payment of fees, and commitments to protect personal data and other requirements, any investigator can access this information. For their part, the investigators have developed codes and packages to allow others to link the curated exposure and confounder data to the Medicare data, and they are prepared to provide the appropriate code and instructions.

The investigators' commitment to making their data and methods publicly available is noteworthy and welcomed; it enables other investigators to access the data, to test different approaches to the analysis, and to move the science forward.

CAUSAL INFERENCE MODELS

In addition to the research discussed earlier, Dominici and colleagues note in their report the importance of, and are devoting significant effort to, the development and extension of methods for causal analysis, an area where they have considerable expertise. This work is increasingly important because of the challenges in accounting for and analyzing all the covariates in the preceding analyses of observational data, and they have made some strides in this direction. The Investigators' Phase 1 report includes only a relatively brief summary of this work — understandably still in progress — so interested readers are advised to go to the referenced papers, which the HEI Review Panel reviewed for details (Wu et al. 2019; Papadogeorgou and Dominici, forthcoming publication; see also Makar et al. 2017). The causal modeling work so far has taken two different directions, described and discussed as follows:

Regression Calibration In the first method, the investigators have developed causal inference approaches based on regression calibration (RC) to account for exposure prediction errors (Wu et al. 2019). A generalized propensity score approach is utilized for confounding adjustment along with the RC to address exposure measurement error. The development of approaches to handle exposure measurement error and confounding in the causal setting would be an important advance given that environmental exposures are almost always prone to error (whether obtained through direct monitoring or via exposure modeling) and confounding bias is a persistent concern in observational studies. Hence, this research is potentially innovative and significant.

However, in its current form this work has several potential limitations that might lower its effectiveness in the setting of ambient air-pollution-related models for which the method is primarily intended. For example:

- It is not immediately clear whether $PM_{2.5}$ concentrations monitored inside a grid cell are error-free exposures for that grid cell, as the investigators assume. Ideally, one would use more flexible methods to allow for the possibility of such errors.
- Given that the internal validation study for the RC step is based on data from monitored locations (likely higher pollution locations compared with nonmonitored locations), it is very likely this sample will be systematically different from the main study sample. Specifically, this situation might violate some of the

assumptions such as "transportability" (i.e., the relationship between true (X) and error-prone (W) exposures, conditional on covariates (D), would be the same in the validation study where X is observed and in the main study in which it is not). The extensive simulation study does not appear to address this issue. Moreover, it doesn't grapple with the complexities of air pollution exposure, the impact of the complicated exposure modeling that produces the exposure estimates and their associated measurement error, or the complicated spatial structures of exposure, outcome, and covariates. This raises questions about the usefulness of this method in the real context of the epidemiological analyses performed in this study.

- The investigators focus on settings for which they have a continuous monitoring data (with error), yet they convert the continuous values into a categorical scale, likely because of technical challenges. It is important that future work attempt to develop similar methods, but for continuous exposure, which is more useful for the ultimate intended application.

Local Exposure–Response Confounding Adjustment In the second method, Dominici and colleagues have developed a new Bayesian causal approach known as local exposure–response confounding adjustment (LERCA), to estimate exposure–response curves accounting for confounding bias under low exposure settings (Papadogeorgou and Dominici, forthcoming publication). This work recognizes and addresses the potentially differential effects of confounders at different levels of exposure and also the model uncertainty associated with confounder selection. The development of an R package to implement the approach, the simulation study to assess performance, and the application to a large data set are some of the notable strengths.

Developing a preliminary directed acyclic graph would be informative in the design and interpretation of models such as the LERCA model. With that as a starting point — a Bayesian prior in essence — the investigation can use the models to inform our understanding of these relationships and modify the underlying conceptual model in what will likely be an ongoing, iterative process. The LERCA model has great potential as a useful new statistical tool, but it is not entirely clear what public health concerns about the data motivated the investigators to develop this specific model, and why differential confounding at different levels of exposure would be expected. It seems at least as likely that confounding might differ for different levels of the confounders given that, unlike the presumed effects of $PM_{2.5}$, these are often not directly causal or have nonmonotonic relationships. Housing value, for example, does not directly cause disease or hospitalization and, as a surrogate for other factors with strong regional variation, is likely to have a complex relationship with this outcome. Temperature has a U-shaped relationship with biological stress and its role as a confounder is likely to vary strongly with temperature level.

A common limitation of both these approaches stems from the different spatial refinement of the data, in other words, between ambient air pollution concentration estimates (at 1 km^2, which are then aggregated to the ZIP code level) and data on health and other covariates (available at the ZIP code level). Neither of the new techniques appears to try to deal with this complexity. This continues to raise questions of exposure error and confounding that potentially affect the primary analyses, a limitation which the investigators specifically note as well. As this work proceeds, a clarification and better understanding of these issues and their impact would be important to the successful completion of the full analyses in this project.

Fully exploring and explaining the observed relationships between air pollution and mortality will necessarily be an iterative process, and the Panel was glad to learn that the investigators plan to spend considerable efforts in this direction in their future work. However, although promising, the current state of methods development is only the first step and may not be a match for the complexity in study design (particularly its hybrid nature), exposure measurement error, and modeling structure of the analysis that has been published using traditional regression-based methods. The investigators have also indicated their plan to develop less computationally intensive methods for analyzing the entire air pollution and health database; it will be informative if the causal models can be applied to those large data sets using these more efficient methods. Given that each of these models relies on assumptions (e.g., accurate measurement of confounders and their full and appropriate specification) to make them mathematically tractable, it is important that the potential impact of these assumptions be explicitly and carefully considered in any interpretation of results as these methods are applied to the larger data sets.

CONCLUSIONS OF THE PANEL'S EVALUATION OF THE PHASE 1 INITIAL ANALYSES

Dominici and colleagues have conducted an extensive and innovative set of initial analyses in these extraordinarily large air pollution and health data sets. They have conducted two distinct types of analyses: a cohort-based analysis of long-term exposures and a case–crossover-based analysis of short-term exposures. They report positive

associations of both $PM_{2.5}$ and O_3 with all-cause mortality, with associations extending to the lowest concentrations and with little evidence of a threshold in these initial analyses. These findings met the criteria for statistical significance, although, as noted earlier, it is important to not overinterpret the statistical robustness of results derived from such a very large data set (Meng 2018). To their credit, the investigators also conducted a range of sensitivity analyses, and they also attempted to control for many key potential confounders in their cohort study that were available in the larger data set, as well as in the smaller Medicare Beneficiaries Survey; in all the analyses to date, these further analyses did not meaningfully change the initial findings of associations.

These initial analyses do make a valuable contribution to the literature; however, while these analyses are thorough and extensive, there is still more work to be done to understand fully the importance of the findings. The investigators are well aware of many of the issues brought up in this commentary and acknowledge them, both in the "Introduction" section and in the "Limitations" section of their Investigators' Report. The Panel was also glad to note that the investigators are proceeding, in completing their project for HEI, with additional analyses and are also developing a less computationally intensive analytic approach in the full cohort. As noted in their discussion of limitations, there are several important analyses that will need to be undertaken before firmer conclusions can be drawn from these studies. Key among important further analyses are:

- *Further analyses of measured and unmeasured confounders*: While the investigators applied the data available on confounders and adopted the AG approach, which offered some advantage over the traditional Cox proportional hazards method in addressing some confounding due to time-dependent covariates, significant questions remain. The Panel discussed these in some detail and would like to highlight here some that will need to be further analyzed:
  - *Potential confounding by time trends*: With air pollution and death rates having declined over the course of the cohort analyses, the degree to which potential confounding of the results may have been affected by time was not adequately analyzed in these initial analyses. The investigators have acknowledged this and indicated they plan to further analyze this important question, by conducting sensitivity analyses using a newly developed causal inference approach.
  - *Potential confounding by other pollutants*: Other air pollutants may also confound the estimates of exposure and effects seen in these analyses. The investigators did test the potential influence of O_3 exposure on PM effects — and vice versa — which was an important strength of their work. In addition, they are now developing an exposure model for NO_2 that will allow adjustment for this pollutant in their final models.
  - *Analysis of spatial confounding and geographical patterns:* As the Panel noted earlier and the investigators acknowledge, the current analyses are conducted at a national level, without fully addressing potentially significant geographical variation in air pollution (both concentrations and composition) and the underlying health status (i.e., variability in $PM_{2.5}$ levels and substantial diversity in levels of obesity across different regions).
- *Spatial scales and the hybrid model:* There are several spatial scales of the many variables in both the long-term and short-term analyses, and the resulting complex quasi-ecological (hybrid) nature of these analyses make it difficult to fully understand the implications of these. For example, as the Panel noted earlier — and despite the considerable efforts by the investigators to estimate exposure accurately — there are some potential sources of error that may affect results. These include, though may not be limited to, (1) potential underestimation of rural concentration levels due to the relative paucity of ground monitors for evaluation and training in those areas; and (2) the potential differences between exposures estimated at a 1-km^2 grid but then applied to health data at the ZIP code level. Although it may not be possible to fully eliminate exposure error from an observational study such as this, the investigators will greatly enhance their final efforts by making every effort to quantify these errors and ideally to account for them in the health analyses.
- *Development, testing, and application of causal inference methods in the full population:* As noted earlier, these analyses would benefit from rigorous application of causal inference methods to the full cohort. To their credit, the investigators have taken initial steps toward developing two such methods and continue to work on them. Properly developed and applied, these methods can also address concerns about residual confounding. The Panel has noted some important questions about these and recommends that the methods be fully evaluated and then applied.

The investigators are to be congratulated for a set of extensive and creative analyses conducted in the largest air pollution and health database to date. While initial conclusions may be drawn from these first analyses, the Panel will wait for the planned extensive further analyses

to be completed before reaching full conclusions on the air pollution and public health implications of this important research.

ACKNOWLEDGMENTS

The Review Panel is grateful to Hanna Boogaard for her oversight of the study; to Rashid Shaikh for his assistance in preparing its Commentary; to Carol Moyer for the editing of this Report and its Commentary; and to Hope Green, Fred Howe, Hilary Selby Polk, and Ruth Shaw for their roles in preparing this Research Report for publication.

REFERENCES

Andersen PK, Gill RD. 1982. Cox's regression model for counting processes: A large sample study. Ann Stat 10:1100–1120; doi:10.1214/aos/1176345976.

Bergen S, Sheppard L, Kaufman JD, Szpiro AA. 2016. Multipollutant measurement error in air pollution epidemiology studies arising from predicting exposures with penalized regression splines. J R Stat Soc C 65(5):731–753; doi:10.1111 /rssc.12144.

Centers for Disease Control and Prevention. 2019. Adult Obesity Prevalence Maps. *https://www.cdc.gov/obesity/data/prevalence-maps.html*. Atlanta, GA:Centers for Disease Control and Prevention. Accessed on October 15, 2019.

Crouse D L, Peters PA, van Donkelaar A, Goldberg MS, Villeneuve PJ, Brion O, et al. 2012. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. Environ Health Perspect 120(5):708–714; doi:10.1289/ehp.1104049.

Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, et al. 2017a. Association of short-term exposure to air pollution with mortality in older adults. JAMA 318(24):2446–2456; PMID: 29279932; doi: 10.1001/jama.2017.17923.

Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J. 2016. Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. Environ Sci Technol 50(9):4712–21: PMID: 27023334; doi:10.1021/acs.est.5b06121.

Di Q, Rowland S, Koutrakis P, Schwartz J. 2017b. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. J Air Waste Manag Assoc 67(1):39–52: PMID: 27332675; doi:10.1080 /10962247.2016.1200159.

Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. 2017c. Air pollution and mortality in the Medicare population. N Engl J Med 376(26):2513–2522; PMID: 28657878; doi:10.1056/NEJMoa1702747.

Dockery DW, Pope CA, Xu X, et al. 1993. An association between air pollution and mortality in six U.S. cities. N Engl J Med 329(24):1753–1759; doi:10.1056/NEJM 199312093292401.

Eftim S, Dominici F. 2005. Multisite time-series studies versus cohort studies: Methods, findings, and policy implications. J Toxicol Environ Health A 68(13-14):1191–205. PubMed PMID: 16024498.

Hajat A, Diez-Roux AV, Adar SD, Auchincloss AH, Lovasi GS, O'Neill MS, et al. 2013. Air pollution and individual and neighborhood socioeconomic status: evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). Environ Health Perspect 121(11-12):1325–1233; doi:10.1289 /ehp.1206337.

Hales S, Blakely T, Woodward A. 2012. Air pollution and mortality in New Zealand: Cohort study. J Epidemiol Community Health 66(5):468–473; doi:10.1136/jech.2010 .112490.

Kloog I, Nordio F, Coull BA, Schwartz J. 2012. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal $PM_{2.5}$ exposures in the mid-Atlantic states. Environ Sci Technol 46:11913–11921.

Künzli N, Medina S, Kaiser R, Quénel P, Horak F Jr, Studnicka M. 2001. Assessment of deaths attributable to air pollution: Should we use risk estimates based on time series or on cohort studies? Am J Epidemiol 153(11):1050–1055.

Makar M, Antonelli J, Di Q, Cutler D, Schwartz J, Dominici F. 2017. Estimating the causal effect of low levels of fine particulate matter on hospitalization. Epidemiology 28(5):627–634; doi:10.1097/EDE.0000000000000690.

Meng X-L. 2018. Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. Ann Appl Stat 12(2): 685–726; doi:10.1214/18-AOAS1161SF.

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

Rabl A. 2003. Interpretation of air pollution mortality: Number of deaths or years of life lost? J Air Waste Manag Assoc 53(1):41–50.

Sabath MB, Di Q, Schwartz J, Braun D, Dominici F, Choirat C. 2018. aipred: A flexible R package implementing methods for predicting air pollution. Presented at the 5th IEEE International Conference on Data Science and Advanced Analytics, 1–4 October 2018, Turin, Italy. Available: *arXiv:1805.11534v2 [stat.ML].*

Szpiro AA, Paciorek CJ. 2013. Measurement error in two-stage analyses, with application to air pollution epidemiology. Environmetrics 24(8):501–517; doi:10.1002/env.2233.

U.S. Environmental Protection Agency (U.S. EPA). 2004. Air Quality Criteria for Particulate Matter (Final Report). EPA 600/P-99/002aF-bF, 2004. Washington, DC:U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency (U.S. EPA). 2013. Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants. Available: *www.epa.gov/isa/integrated-science-assessment-isa-ozone-and-related-photochemical-oxidants.*

U.S. Environmental Protection Agency (U.S. EPA). 2016. Particulate Matter (PM) Standards — Table of Historical PM NAAQS. Available: *www3.epa.gov/ttn/naaqs/standards/pm/s_pm_history.html* [accessed 4 October 2019].

U.S. Environmental Protection Agency (U.S. EPA). Particulate Matter ($PM_{2.5}$) Trends. Available: *www.epa.gov/air-trends/particulate-matter-pm25-trends* [accessed 24 April 2019].

U.S. Environmental Protection Agency (U.S. EPA). 2015. Preamble to the Integrated Science Assessments (ISA). EPA/600/R-15/067. Washington, DC: U.S. Environmental Protection Agency.

Wu X, Braun D, Kioumourtzoglou MA, Choirat C, Di Q, Dominici F. 2019. Causal inference in the context of an error prone exposure: Air pollution and mortality. Annals of Applied Statistics 13(1):520–547.

ABBREVIATIONS AND OTHER TERMS

AG	Andersen Gill
AOD	aerosol optical depth
AQS	air quality system
BMI	body mass index
CASAC	Clean Air Scientific Advisory Committee
CI	confidence interval
CMS	Centers for Medicare and Medicaid Services
CTM	chemical transport model
df	degrees of freedom
ER	exposure–response
GEE	generalized estimating equation
GEOS-Chem	Goddard Earth Observing System (chemical transport model)
GPS	generalized propensity score
IPTW	inverse probability treatment weighting
LERCA	local exposure–response confounding adjustment
MCBS	Medicare Current Beneficiary Survey
MODIS	moderate resolution imaging spectro-radiometer
NAAQS	National Ambient Air Quality Standards
NASA	National Aeronautics and Space Administration
NDVI	normalized difference vegetation index
NO_2	nitrogen dioxide
NOAA	National Oceanic and Atmospheric Administration (U.S.)
NO_x	nitrogen oxides
O_3	ozone
OMI	ozone monitoring instrument
PM	particulate matter
$PM_{2.5}$	particulate matter ≤ 2.5 μm in aerodynamic diameter
RC	regression calibration
RC-GPS	regression calibration-generalized propensity score
ResDAC	Research Data Assistance Center
RR	relative risk
RRI	relative risk increase
SES	socioeconomic status
U.S. CAA	U.S. Clean Air Act
U.S. EPA	U.S. Environmental Protection Agency
VOC	volatile organic compound

HEI BOARD, COMMITTEES, and STAFF

Board of Directors

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

Enriqueta Bond *President Emerita, Burroughs Wellcome Fund*

Jo Ivey Boufford *President, International Society for Urban Health*

Homer Boushey *Emeritus Professor of Medicine, University of California, San Francisco*

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*

Martha J. Crawford *Dean, Jack Welch College of Business and Technology, Sacred Heart University*

Michael J. Klag *Dean Emeritus and Second Century Distinguished Professor, Johns Hopkins Bloomberg School of Public Health*

Alan I. Leshner *CEO Emeritus, American Association for the Advancement of Science*

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

Research Committee

David A. Savitz, Chair *Professor of Epidemiology, School of Public Health, and Professor of Obstetrics and Gynecology, Alpert Medical School, Brown University*

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

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*

Amy H. Herring *Sara & Charles Ayres Professor of Statistical Science and Global Health, Duke University, Durham, North Carolina*

Barbara Hoffmann *Professor of Environmental Epidemiology, Institute of Occupational, Social, and Environmental 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*

Review Committee

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

Kiros Berhane *Professor of Biostatistics and Director of Graduate Programs in Biostatistics and Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California*

Michael Jerrett *Professor and Chair, Department of Environmental Health Sciences, Fielding School of Public Health, University of California, Los Angeles*

Frank Kelly *Professor of Environmental Health and Director of the Environmental Research Group, King's College London*

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

Jennifer L. Peel *Professor of Epidemiology, Colorado School of Public Health and Department of Environmental and Radiological Health Sciences, Colorado State University*

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

HEI BOARD, COMMITTEES, and STAFF

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*
Emily Alden *Corporate Secretary*

Lee Ann Adelsheim *Research Assistant*
Hanna Boogaard *Consulting Senior Scientist*
Sofia Chang-DePuy *Digital Communications Manager*
Aaron J. Cohen *Consulting Principal Scientist*
Robert M. Davidson *Staff Accountant*
Philip J. DeMarco *Compliance Manager*
Hope Green *Editorial Project Manager*
Joanna Keel *Research Assistant*
Lissa McBurney *Science Administrative Assistant*
Pallavi Pant *Staff Scientist*
Allison P. Patton *Staff Scientist*
Hilary Selby Polk *Managing Editor*
Anna S. Rosofsky *Staff Scientist*
Robert A. Shavers *Operations Manager*
Annemoon M.M. van Erp *Managing Scientist*
Eleanne van Vliet *Staff Scientist*
Donna J. Vorhees *Director of Energy Research*
Katherine Walker *Principal Scientist*

HEALTH
EFFECTS
INSTITUTE

75 Federal Street, Suite 1400
Boston, MA 02110, USA
+1-617-488-2300
www.healtheffects.org

RESEARCH
REPORT

Number 200
November 2019