



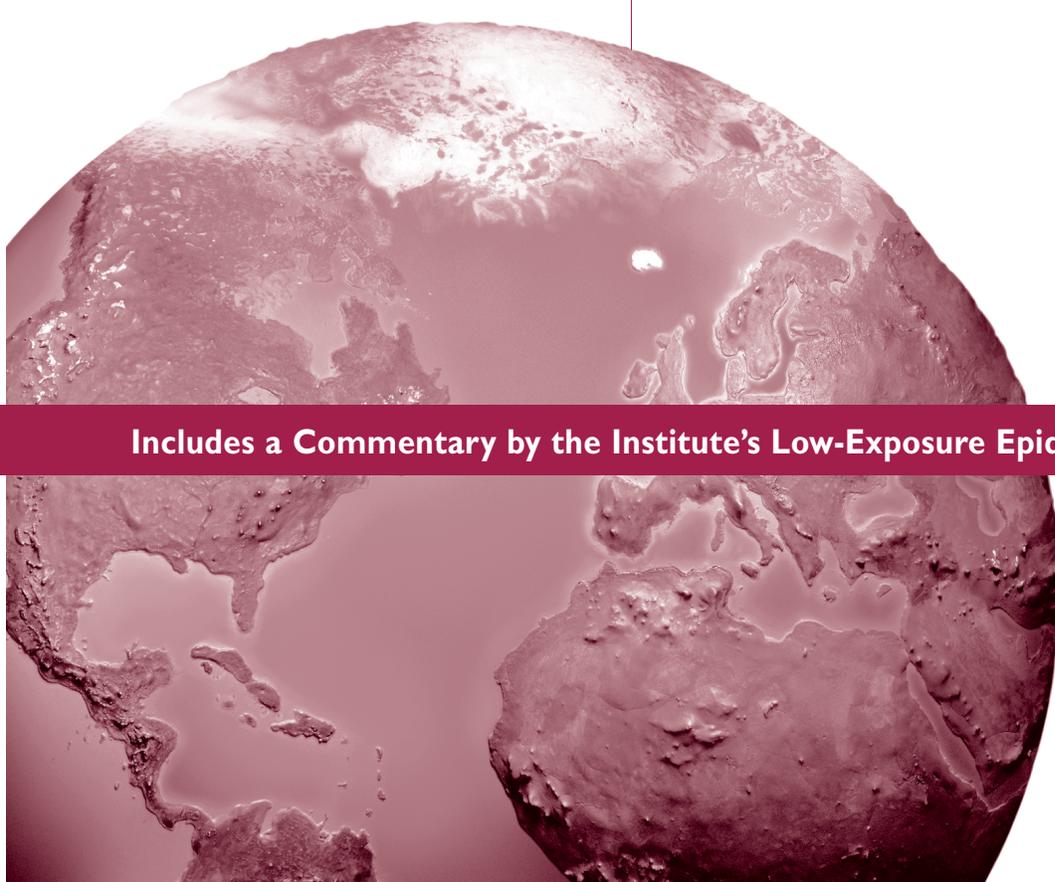
RESEARCH REPORT

HEALTH
EFFECTS
INSTITUTE

Number 211
January 2022

Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Implementation of Causal Inference Methods

Francesca Dominici, Antonella Zanobetti, Joel Schwartz,
Danielle Braun, Ben Sabath, and Xiao Wu



Includes a Commentary by the Institute'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: Implementation of Causal Inference Methods

Francesca Dominici, Antonella Zanobetti, Joel Schwartz,
Danielle Braun, Ben Sabath, and Xiao Wu

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

Research Report 211
Health Effects Institute
Boston, Massachusetts

Trusted Science · Cleaner Air · Better Health

Publishing history: This document was posted at www.healtheffects.org in January 2022.

Citation for document:

Citation for document: Dominici F, Zanobetti A, Schwartz J, Braun D, Sabath B, Wu X. 2022. Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Implementation of Causal Inference Methods. Research Report 211. Boston, MA: Health Effects Institute.

© 2022 Health Effects Institute, Boston, Mass., U.S.A. David Wade, Virginia Beach, Va., Graphic Designer. 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	vii
About This Report	ix
Contributors	xi
Preface	xiii
HEI STATEMENT	I
INVESTIGATORS' REPORT <i>by Dominici et al.</i>	5
ABSTRACT	5
I. INTRODUCTION	6
1.1. Background and Motivation	6
1.2. Specific Aims of the Research Proposal	6
1.3. Executive Summary of the Phase I Report	7
1.4. Challenges Identified in the Phase I Report	9
1.5. Harmonization of the Phase I and Final Reports	9
1.6. Roadmap to the Content of This Final Report	10
2. EXPOSURE ASSESSMENT FOR PM_{2.5}, NO₂, AND O₃	11
2.1. Exposure Assessment for PM _{2.5}	11
2.2. Exposure Assessment for NO ₂	13
2.3. Exposure Assessment for O ₃	13
3. NATIONAL CAUSAL INFERENCE ANALYSIS ON LONG-TERM EFFECTS OF PM_{2.5} ON ALL-CAUSE MORTALITY	13
3.1. Why Do We Need Causal Inference Methods in Addition to Standard Regression Approaches?	13
3.2. Data	16
3.2.1. Exposure Assessment	16
3.2.2. Potential Confounders	16
3.2.3. Data Linkage	16
3.3. Methods	16
3.3.1. Study Population	16
3.3.2. Statistical Analysis	18
3.4. Results	21
4. NATIONAL CAUSAL INFERENCE ANALYSIS ON LONG-TERM EXPOSURE-RESPONSE CURVES FOR PM_{2.5}, NO₂, AND O₃ ON ALL-CAUSE MORTALITY	24
5. PIPELINE FOR REPRODUCIBLE RESEARCH	27
5.1. Introduction	27
5.2. Replicability Versus Reproducibility	27
5.3. Toward Open Science in Air Pollution Epidemiology: A First	27
5.4. Details on the Pipeline for Ensuring Reproducibility	28
5.4.1. Computational Aspects	30

Research Report 211

6. ONGOING WORK	30
6.1. Harmonized Analyses Across U.S., Canadian, and European Cohorts	30
6.2. Software Development	30
7. DISCUSSION AND CONCLUSIONS	30
ACKNOWLEDGMENTS	33
REFERENCES	33
HEI QUALITY ASSURANCE STATEMENT	36
MATERIALS AVAILABLE ON THE HEI WEBSITE	37
ABOUT THE AUTHORS	38
OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH	39
COMMENTARY by HEI's Low-Exposure Epidemiology Studies Review Panel	41
INTRODUCTION	41
SCIENTIFIC AND REGULATORY BACKGROUND	41
Setting Air Quality Standards in the United States	42
Evaluating Associations Below Current Air Quality Standards and Guidelines	42
SUMMARY OF APPROACH AND METHODS	43
Study Objectives	43
Methods and Study Design	43
Exposure Modeling	43
Study Population	45
Exposure Assignment	45
Main Epidemiological Analyses	45
Additional Epidemiological Analyses	46
SUMMARY OF KEY FINDINGS	47
Modeling and Exposure Estimation Results	47
Epidemiological Results	47
Exposure–Response Functions	47
Reproducible Research	48
EVALUATION BY THE HEI LOW-EXPOSURE EPIDEMIOLOGY STUDIES REVIEW PANEL	48
Evaluation of Study Design and Approach	50
Air Pollution Models and Exposure Estimation	50
Evaluation of Epidemiological Analysis	50
Discussion of the Findings and Interpretation	51
Conclusions	52
ACKNOWLEDGMENTS	52
REFERENCES	52

Research Report 211

Related HEI Publications	55
Abbreviations and Other Terms	56
HEI Board, Committees, and Staff	57

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 2,500 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 Review 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 211, *Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Implementation of Causal Inference Methods*, presents a research project 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 HEI 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 the HEI 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 is first examined by outside technical reviewers and a biostatistician. The report and the reviewers' comments are then evaluated by members of an independent Panel of distinguished scientists who are not involved in selecting or overseeing HEI studies. During the review process, the investigators have an opportunity to exchange comments with the 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 *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*

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

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*

HEI PROJECT STAFF

Hanna Boogaard *Consulting Principal Scientist (Study Oversight)*

Dan Crouse *Senior Scientist (Report Review and Commentary)*

Martha Ondras *Research Fellow (Report Review and Commentary)*

Eva Tanner *Staff Scientist (Report Review)*

George Simonson *Consulting Editor*

Kristin Eckles *Senior Editorial Manager*

Hope Green *Editorial Project Manager*

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 few decades in North America, Europe, and in other developed regions. Despite the decreasing levels of air pollution, several large epidemiological studies published in the early 2010s reported associations between adverse health effects and exposure to air pollution. These studies found associations between exposure to fine particulate matter (PM_{2.5}*) and mortality at levels below the then-current ambient air quality standards (e.g., Beelen et al. 2014a, b; Crouse et al. 2012; Hales et al. 2012; Preface Figure 1). In order to inform future risk assessment and regulation, it is important to confirm whether associations with adverse health effects continue to be observed as levels of air pollution decline still further. It is also important to better understand the shape of the exposure–response function at those low levels. Both issues remain major uncertainties in setting air quality standards.

The growing scientific evidence for effects at pollution levels below current air quality standards, the large overall estimates of the burden of disease attributable to air pollution, 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 was 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 (O₃), 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.

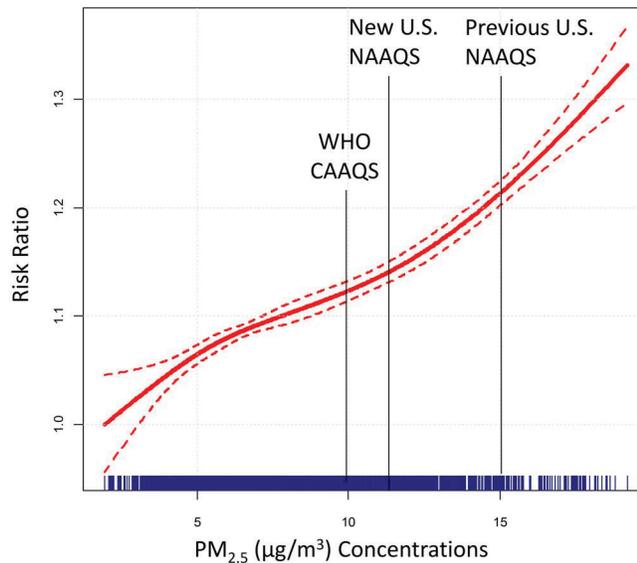
In 2014, HEI issued RFA 14-3, *Assessing Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution*, to solicit studies to address these important questions. The main goals of the RFA were to (1) fund studies to assess health effects of long-term exposure to low levels of ambient air 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 and may also address related questions about health effects at low levels of ambient air pollution; and (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.

Investigators 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 to developing exposure assessment approaches and statistical methodology to enable a robust examination of the associations.

Specifically, investigators were asked to propose studies to:

1. Compare and contrast alternative analytic models and their uncertainty. For example, compare threshold against nonthreshold models, linear against nonlinear models, and parametric against nonparametric models, to characterize

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



Preface Figure 1. Shape of concentration–response function for mortality associated with fine particulate matter in a Canadian cohort. (Adapted from Crouse et al. 2012, courtesy R. Burnett.)

- the exposure–response function(s) at low levels of ambient air pollution.
2. Explore possible variability in estimates of risk at low levels among populations and identify possible contributing factors. Such factors could include age, smoking, socio-economic 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 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.

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

Preface

OVERVIEW OF THE 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 the 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 1). The studies included large population cohorts (with detailed individual information about potential confounders on all or a subset of the cohort) as well as large administrative databases with greater statistical power (albeit with less individual covariate information). Additionally, the three teams employed satellite data and ground-level exposure measurements, used high-quality exposure assessment models at high spatial resolutions, and set out to develop and apply novel statistical methods.

The three studies were expected to inform the scientific community and risk assessors and policy makers regarding exposure–response functions at levels of

ambient air pollution currently prevalent in North America, Western Europe, and other developed regions. The full sets of analyses were expected to be completed in 2021 (see below).

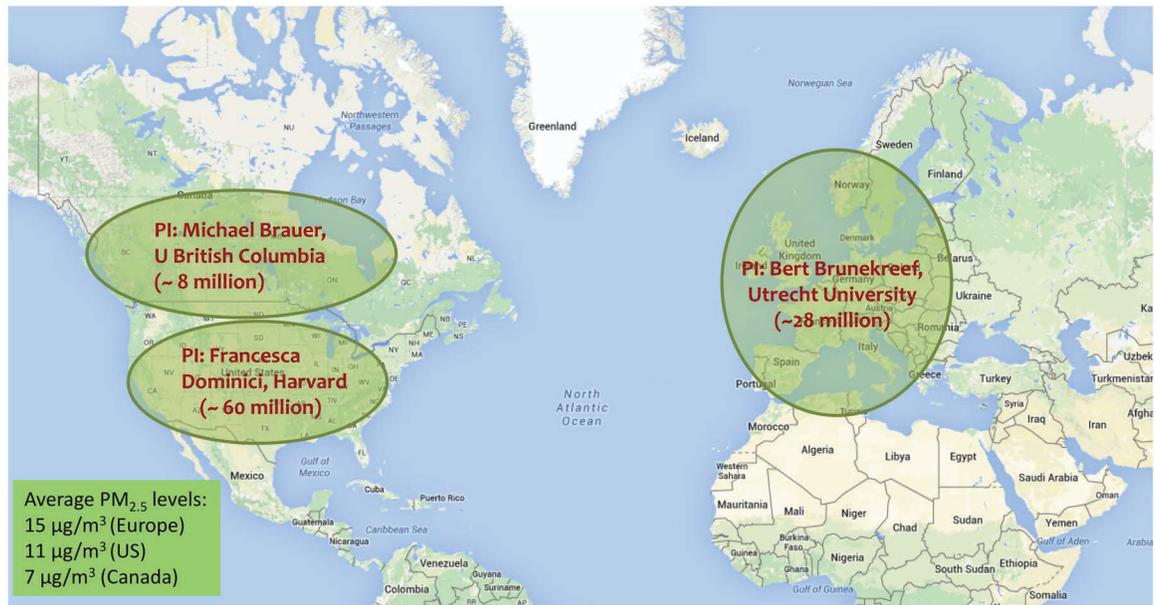
CANADIAN STUDY (MICHAEL BRAUER ET AL.)

Brauer and colleagues proposed to assess 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 used Canadian census data and had access to a nationally representative population of approximately 8.5 million Canadians (ages 25–90 yr) (Preface Figure 2). The Canadian team proposed developing hybrid models using primarily satellite data, as well as chemical transport models, land use variables, and routinely collected monitoring data for $PM_{2.5}$, as well as estimating exposures for NO_2 and O_3 for Canada and the United States during the period 1981–2016. Additionally, they planned to validate satellite data against ground-based monitors in Canada as part of the SPARTAN network (Snider et al. 2015).

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

Investigator (institution)	Study Title	Phase 1 Report	Final Report Published
Brauer, Michael (University of British Columbia, Canada)	Mortality–Air Pollution Associations in Low Exposure Environments (MAPLE)	HEI Research Report 203 (2019)	HEI Research Report 212 (2022; in review)
Brunekreef, Bert (Utrecht University, the Netherlands)	Mortality and Morbidity Effects of Long-Term Exposure to Low-Level $PM_{2.5}$, BC , NO_2 , and O_3 : An Analysis of European Cohorts in the ELAPSE Project	None	HEI Research Report 208 (September 2021)
Dominici, Francesca (Harvard T.H. Chan School of Public Health, USA)	Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution	HEI Research Report 200 (2019)	HEI Research Report 211 (January 2022)

Preface



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

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

1. About 2.6 million subjects who completed the 1991 census long-form of the Canadian Census Health and Environment Cohorts (CanCHEC),
2. About 3.5 million subjects who completed the 1996 CanCHEC census long-form,
3. About 3.5 million subjects who completed the 2001 CanCHEC census long-form, and
4. About 540,000 subjects 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 based 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. 2014a, b; Cesaroni et al. 2014; Eeftens et al. 2012a, b). In the

current HEI-funded study, the investigators proposed to analyze pooled data from 10 ESCAPE cohorts (instead of the cohort-specific approach they used previously). In addition, they planned to use data from six large administrative cohorts to yield a total study population of approximately 28 million Europeans (Preface Figure 2). They proposed developing hybrid Europewide and location-specific exposure models that would utilize land use information, dispersion modeling, satellite data, ESCAPE monitoring data, and routinely collected monitoring data for PM_{2.5}, NO₂, O₃, 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 proposed to investigate 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 into the merits of both approaches.

Preface

UNITED STATES STUDY (FRANCESCA DOMINICI ET AL.)

Dominici and colleagues proposed to evaluate Medicare and Medicaid data for a study population of approximately 60 million Americans (Preface Figure 2). They planned to develop hybrid exposure models that incorporate satellite data, chemical transport models, land use, and weather variables, and routinely collected monitoring data for PM_{2.5} and its components, NO₂, and O₃, at high spatial resolution (1-km² grid) for the continental United States during the period 2000–2014. Exposure models were to be applied to estimate adverse health effects of air pollution in three cohorts:

1. Medicare enrollees (28.6 million elderly enrollees per year, 2000–2014);
2. Medicaid enrollees (28 million enrollees per year, 2010–2014); 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 planned to analyze the following health outcomes: time to hospitalization by cause, disease progression (time to rehospitalization), and time to death. They proposed developing and applying new causal inference methods to estimate exposure–response functions to adjust for confounding and exposure measurement error. Additionally, they proposed developing tools for reproducible research including approaches for data sharing, record linkage, and statistical software.

STUDY OVERSIGHT

HEI's independent Low Exposure Epidemiology Oversight Panel provided advice and feedback on the study designs, analytical plans, and study progress throughout the duration of the research program (see Contributors' page).

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 worked actively 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 regularly held investigator workshops and site visits, among other activities. In addition, the studies were subject to HEI's special quality assurance procedures that included an audit by an independent audit team (see www.healtheffects.org/research/quality-assurance).

REVIEW OF PHASE I AND FINAL (PHASE 2) REPORTS

To inform the ongoing review of the U.S. National Ambient Air Quality Standards (NAAQS) for PM_{2.5} and O₃ during 2019–2020, HEI requested Phase I reports based on the research completed during the first two years of the Canadian and U.S. studies. Thus, the Phase I reports by Drs. Brauer and Dominici provided summaries of results to date, including those published in journal articles, with accompanying Commentaries by an independent Special Review Panel. These Phase I reports provided an opportunity to review the results to date and evaluate their strengths and weaknesses, a process normally performed after a study has been completed.

As is common for major research programs, HEI convened a Special Review Panel to independently review the Phase I reports by Drs. Brauer and Dominici. The Panel consists of seven experts in epidemiology, exposure assessment, and biostatistics (see Contributors' page). The Panel also reviewed the final (Phase 2) reports of the three studies.

The three studies commenced in Spring 2016 and were completed at different times in 2020, with final reports published during 2021. In addition, further analyses, for example to compare approaches among the three teams, are ongoing and are expected to be completed at the end of 2021.

REFERENCES

Beelen R, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, Hoffmann B, et al. 2014a. 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.

Preface

Beelen R, Stafoggia M, Raaschou-Nielsen O, Andersen ZJ, Xun WW, Katsouyanni K, et al. 2014b. Long-term exposure to air pollution and cardiovascular mortality: an analysis of 22 European cohorts. *Epidemiology* 25:368–378; doi: 10.1097/EDE.0000000000000076.

Cesaroni G, Forastiere F, Stafoggia M, Andersen ZJ, Badaloni C, Beelen R, et al. 2014. Long term exposure to ambient air pollution and incidence of acute coronary events: Prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE project. *BMJ* 348:f7412; doi: 10.1136/bmj.f7412.

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.

Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, et al. 2012a. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10}

and PM_{coarse} in 20 European study areas: Results of the ESCAPE project. *Environ Sci Technol* 46:11195–11205; doi: 10.1021/es301948k.

Eeftens M, Tsai MY, Ampe C, Anwander B, Beelen R, Bellander T, et al. 2012b. Spatial variation of $PM_{2.5}$, PM_{10} , $PM_{2.5}$ absorbance and PM_{coarse} concentrations between and within 20 European study areas and the relationship with NO_2 : Results of the ESCAPE project. *Atmos Environ* 62:303–317.

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

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* 8:505–521; doi.org/10.5194/amt-8-505-2015.

HEI STATEMENT

Synopsis of Research Report 211

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

BACKGROUND

The growing scientific evidence reporting effects of air pollution on health at concentrations below current air quality standards and the large burden of disease attributed to air pollution suggest that more stringent air quality standards and guidelines will likely be considered in the future. To improve our understanding of exposure–response functions for mortality and morbidity at low concentrations of PM_{2.5}, NO₂, O₃, and other ambient air pollutants, HEI issued RFA 14-3, *Assessing Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution*. Three studies based in the United States, Canada, and Europe were funded that used state-of-the-art exposure methods and large cohorts in high-income countries where ambient concentrations are generally low (i.e., lower than current air quality guidelines and standards for Europe and the United States). HEI convened an independent Low-Exposure Epidemiology Studies Review Panel to evaluate the studies’ strengths and weaknesses. This Statement highlights results from the study in the United States.

APPROACH

Dominici and colleagues aimed to address some of the knowledge gaps related to health effects of long-term exposures to low concentrations of air pollution using a cohort of 68.5 million older Americans enrolled in the U.S. Medicare program. Their approach included modeling spatial and temporal patterns of ambient air pollution, developing cutting-edge causal inference statistical models, describing risks to mortality associated with exposures in a very large dataset, and making the methods and data available to the wider scientific community. The study had four broad aims:

1. **Exposure Prediction and Data Linkage**
Estimate long-term exposures to concentrations of ambient PM_{2.5}, O₃, and NO₂ at

What This Study Adds

- This study evaluated the risk of mortality associated with exposure to low ambient air pollution concentrations in a cohort of 68.5 million older Americans.
- The investigators developed annual exposure models for fine particulate matter (PM_{2.5}), nitrogen dioxide (NO₂), and ozone (O₃) at a spatial resolution of 1 km × 1 km for the years 2000 to 2016 covering the contiguous United States.
- They presented results from three newly developed causal inference approaches and from two traditional regression approaches.
- The investigators reported increased risks of all-cause mortality of 6% to 8% per 10-µg/m³ increase in PM_{2.5} across the five approaches, with larger effect estimates in a low exposure subcohort.
- The consistency of the associations across methods provides stronger support than past studies for what is likely a causal effect between long-term exposure to PM_{2.5} and mortality.

1 km × 1 km spatial resolution for the contiguous United States in 2000–2016.

2. **Causal Inference Methods for Exposure–Response Functions** Develop a new causal inference framework to estimate a nonlinear exposure–response function, adjust for measured and unmeasured confounders, and detect effect modification in the presence of multiple exposures.
3. **Evidence of Adverse Health Effects** Apply methods developed in Aim 2, along with traditional regression approaches, to estimate all-cause mortality by year and zip code associated with long-term exposure to ambient air pollution for cohort participants. Individual-level data included 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.

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 211 contains both the detailed Investigators’ Report and a Commentary on the study prepared by the Institute’s Low-Exposure Epidemiology Studies Review Panel.

4. Tools for Data Access and Reproducibility

Develop approaches for data sharing, record linkage, and statistical software to foster transparency and reproducibility of the work.

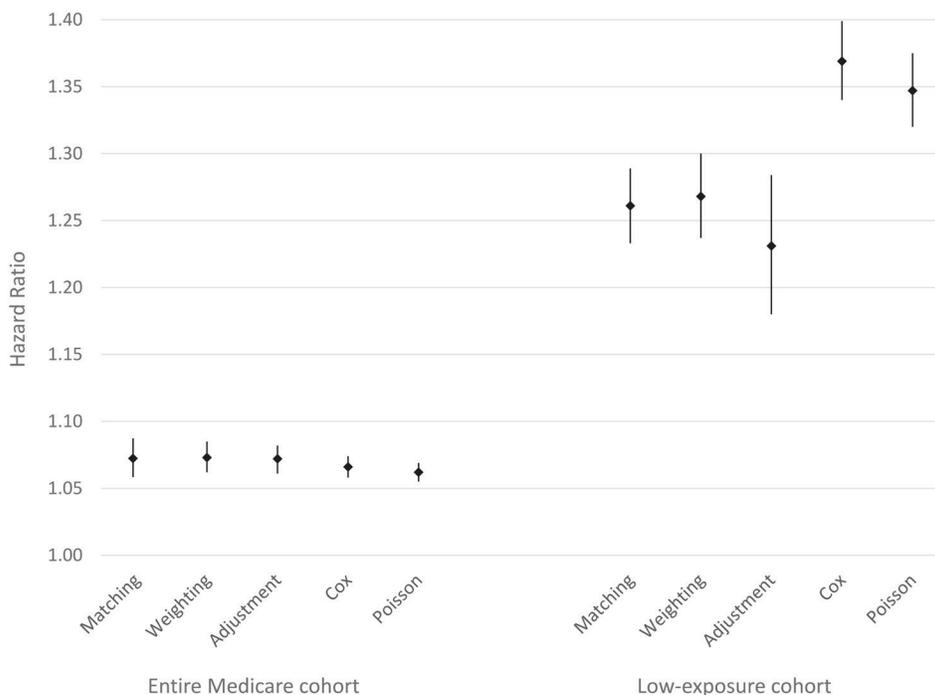
The exposure model inputs included monitoring data from the U.S. EPA Air Quality System, satellite-derived aerosol optical depth, meteorological variables, land-use variables that represent local emissions and small-scale variations in concentrations (e.g., road density, elevation, and normalized difference vegetation index), and daily predictions from two chemical-transport models to simulate atmospheric components. The investigators assigned the predicted annual average exposures to cohort participants' residential zip code for each year of follow-up.

The causal inference approach used generalized propensity scores and attempted to mimic a randomized study. They applied three causal modeling approaches, namely matching, weighting, and adjustment. The propensity scores were estimated by modeling zip code-level exposures conditional on area-level risk factors, meteorological variables, and year and region. They also applied two traditional regression approaches, namely Cox and Poisson models. Throughout the study, they examined health effects for the entire cohort and for a

subpopulation exposed to annual average $PM_{2.5}$ concentrations below or equal to $12 \mu\text{g}/\text{m}^3$ during every year of follow-up (i.e., low exposure subcohort) in order to address the question of health effects below the current U.S. annual National Ambient Air Quality Standard for $PM_{2.5}$. They also performed additional analyses, using, for example, single- and multipollutant models.

KEY RESULTS

The investigators estimated that the mean $PM_{2.5}$ exposure for cohort participants was $9.8 \mu\text{g}/\text{m}^3$, well below the current U.S. standard of $12 \mu\text{g}/\text{m}^3$. The investigators reported consistent, statistically significant results for the five statistical approaches. Specifically, hazard ratios and 95% confidence intervals associated with a $10\text{-}\mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ exposure were 1.07 (1.06, 1.07) for the Cox regression, 1.06 (1.06, 1.07) for the Poisson regression, 1.07 (1.05, 1.08) for general propensity score matching, 1.08 (1.07, 1.09) for score weighting, and 1.07 (1.06, 1.08) for score adjustment (see Statement Figure). The investigators found notably larger effect estimates with the low-exposure subcohort. For example, with Cox regression, they reported a hazard ratio of 1.37 (1.34, 1.40).



Statement Figure. Associations between longer-term exposures to $PM_{2.5}$ and all-cause mortality among enrollees in the full Medicare cohort (left side) and low-exposure cohort (right side). Hazard ratios (calculated per $10\text{-}\mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ exposure) and 95% confidence intervals were estimated using three causal inference approaches with generalized propensity scores (matching, weighting, and adjustment) and two traditional approaches (Cox and Poisson regression). (Source: Investigators' Report Figure 6).

In single-pollutant models, the investigators found evidence of increased risk of mortality associated with long-term $PM_{2.5}$ exposures across the range of annual average $PM_{2.5}$ concentrations between 2.8 and 17.2 $\mu\text{g}/\text{m}^3$, which included 98% of observations. The exposure–response function for $PM_{2.5}$ was almost linear at exposures below the current U.S. standard. They found evidence of a relationship between mortality and long-term NO_2 exposures at higher concentrations (associations at exposures below annual mean ≤ 53 ppb [approximately 100 $\mu\text{g}/\text{m}^3$] were nonlinear and statistically uncertain). Similarly, the exposure–response function for long-term O_3 exposures and mortality showed some evidence of increased risks at exposures higher than 45 ppb (approximately 88 $\mu\text{g}/\text{m}^3$), but the exposure–response function was almost flat at concentrations below that, showing no statistically significant effect. Generally, adjusting for the other two pollutants slightly attenuated the effects of $PM_{2.5}$ on mortality and slightly elevated the effects of NO_2 exposure; results for O_3 remained almost unchanged.

INTERPRETATION AND CONCLUSIONS

The HEI Low-Exposure Epidemiology Studies Review Panel concluded that this report presents a high-quality and thorough investigation into associations between risk of mortality and exposures to ambient air pollution in the United States. The finding of increased risks of all-cause mortality in the low exposure subcohort across the various analytical approaches increases the confidence that mortality is associated with long-term concentrations of $PM_{2.5}$ below the current U.S. standard. The investigators also reported adverse associations between O_3 and NO_2 with mortality, but not at the lowest concentrations.

The stronger effects reported in the low-exposure subgroup could be due in part to those in that group being more susceptible to the effects of exposure. For example, the low-exposure subcohort excluded participants in large areas of the Eastern United States

and likely excluded most people in most major cities. Whereas the main analyses describe the risk for the elderly U.S. population as a whole, the low-exposure analyses to some extent describe the risk for those in smaller towns and rural areas (who tend to be of lower socioeconomic status, have poorer health behaviors, more limited access to health services, and have a higher prevalence of diabetes or other comorbidities that might also increase susceptibility to the effects of exposure).

Particularly strong aspects of this work included the use of an extremely large, national health cohort; relatively high-resolution annual mean exposure estimates for each year of follow-up; the development of novel approaches to causal modeling to assess the associations between air pollution exposure and mortality; and comparisons of results from these with results from traditional approaches. The evaluation of the nonlinearity in multipollutant models was an additional valuable contribution. The Panel appreciated that datasets and statistical codes have been made publicly available, thus facilitating transparency and reproducibility.

The Panel had concerns, however, about some of the approaches used, such as the quality of the exposure estimates in rural areas; the fact that all exposure estimates were aggregated to the zip code level of analysis; and the hybrid nature of the study design, which included covariates measured variously at the individual, zip code, and county level.

Ultimately, the major contribution of this study is that it found associations between exposure to low concentrations of $PM_{2.5}$ and mortality in a large cohort of older Americans, with larger effects at the lowest levels of exposure. The fact that the study produced findings using several different causal inference approaches that were generally consistent with each other and with those of previous studies strengthens confidence in the results.

Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Implementation of Causal Inference Methods

Francesca Dominici¹, Antonella Zanobetti¹, Joel Schwartz¹, Danielle Braun^{1,2}, Ben Sabath¹, Xiao Wu¹

¹Harvard T.H. Chan School of Public Health, Boston, Massachusetts; ²Department of Data Sciences, Dana-Farber Cancer Institute, Boston, Massachusetts

ABSTRACT

This report provides a final summary of the principal findings and key conclusions of a study supported by an HEI grant aimed at “Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution.” It is the second and final report on this topic. The study was designed to advance four critical areas of inquiry and methods development. First, it focused on predicting short- and long-term exposures to ambient fine particulate matter (PM_{2.5}*), nitrogen dioxide (NO₂), and ozone (O₃) at high spatial resolution (1 km × 1 km) for the continental United States over the period 2000–2016 and linking these predictions to health data. Second, it developed new causal inference methods for estimating exposure–response (ER) curves (ERCs) and adjusting for measured confounders. Third, it applied these methods to claims data from Medicare and Medicaid beneficiaries to estimate health effects associated with short- and long-term exposure to low levels of ambient air pollution. Finally, it developed pipelines for reproducible research, including approaches for data sharing, record linkage, and statistical software. Our HEI-funded work has supported an extensive portfolio of analyses and the development of statistical methods that can be used to robustly understand the health effects of short- and long-term exposure to low levels of ambient air pollution. Our Phase 1 report (Dominici et al. 2019) provided a high-level overview of our statistical methods, data analysis, and key findings, grouped into the following five areas: (1) exposure prediction, (2) epidemiological studies of ambient

exposures to air pollution at low levels, (3) sensitivity analysis, (4) methodological contributions in causal inference, and (5) an open access research data platform. The current, final report includes a comprehensive overview of the entire research project.

Considering our (1) massive study population, (2) numerous sensitivity analyses, and (3) transparent assessment of covariate balance indicating the quality of causal inference for simulating randomized experiments, we conclude that conditionally on the required assumptions for causal inference, our results collectively indicate that long-term PM_{2.5} exposure is likely to be causally related to mortality. This conclusion assumes that the causal inference assumptions hold and, more specifically, that we accounted adequately for confounding bias. We explored various modeling approaches, conducted extensive sensitivity analyses, and found that our results were robust across approaches and models. This work relied on publicly available data, and we have provided code that allows for reproducibility of our analyses.

Our work provides comprehensive evidence of associations between exposures to PM_{2.5}, NO₂, and O₃ and various health outcomes. In the current report, we report more specific results on the causal link between long-term exposure to PM_{2.5} and mortality, even at PM_{2.5} levels below or equal to 12 µg/m³ and mortality among Medicare beneficiaries (ages 65 and older). This work relies on newly developed causal inference methods for continuous exposure.

For the period 2000–2016, we found that all statistical approaches led to consistent results: a 10-µg/m³ decrease in PM_{2.5} led to a statistically significant decrease in mortality rate ranging between 6% and 7% (= 1 – 1/hazard ratio [HR]) (HR estimates 1.06 [95% CI, 1.05 to 1.08] to 1.08 [95% CI, 1.07 to 1.09]). The estimated HRs were larger when studying the cohort of Medicare beneficiaries that were always exposed to PM_{2.5} levels lower than 12 µg/m³ (1.23 [95% CI, 1.18 to 1.28] to 1.37 [95% CI, 1.34 to 1.40]).

Comparing the results from multiple and single pollutant models, we found that adjusting for the other two pollutants slightly attenuated the causal effects of PM_{2.5} and slightly elevated the causal effects of NO₂ exposure on all-cause mortality. The results for O₃ remained almost unchanged.

This Investigators' Report is one part of Health Effects Institute Research Report 211, which also includes a Commentary by the HEI 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. No potential conflict of interest was reported by the authors.

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.

We found evidence of a harmful causal relationship between mortality and long-term $PM_{2.5}$ exposures adjusted for NO_2 and O_3 across the range of annual averages between 2.77 and 17.16 $\mu g/m^3$ (included >98% of observations) in the entire cohort of Medicare beneficiaries across the continental United States from 2000 to 2016. Our results are consistent with recent epidemiological studies reporting a strong association between long-term exposure to $PM_{2.5}$ and adverse health outcomes at low exposure levels. Importantly, the curve was almost linear at exposure levels lower than the current national standards, indicating aggravated harmful effects at exposure levels even below these standards.

There is, in general, a harmful causal impact of long-term NO_2 exposures to mortality adjusted for $PM_{2.5}$ and O_3 across the range of annual averages between 3.4 and 80 ppb (included >98% of observations). Yet within low levels (annual mean ≤ 53 ppb) below the current national standards, the causal impacts of NO_2 exposures on all-cause mortality are nonlinear with statistical uncertainty.

The ERCs of long-term O_3 exposures on all-cause mortality adjusted for $PM_{2.5}$ and NO_2 are almost flat below 45 ppb, which shows no statistically significant effect. Yet we observed an increased hazard when the O_3 exposures were higher than 45 ppb, and the HR was approximately 1.10 when comparing Medicare beneficiaries with annual mean O_3 exposures of 50 ppb versus those with 30 ppb.

1. INTRODUCTION

1.1. BACKGROUND AND MOTIVATION

The United States Environmental Protection Agency (U.S. EPA) relies on Regulatory Impact Analyses (Stodden et al.) to shape the development of regulatory policies (U.S. Environmental Protection Agency 2020a). Regulatory Impact Analyses have historically relied on ERCs from epidemiological studies, such as the Harvard Six Cities study (Dockery et al. 1993) to estimate the health events that would be prevented by regulation-induced reductions in pollution exposures. However, given the increased interest in the development of methods for causal inference and their potential value in regulatory decisions, new statistical methods and new nationwide epidemiological studies are needed to estimate ERCs that are grounded in a causal inference framework to rigorously adjust for confounding and reduce modeling assumptions.

1.2. SPECIFIC AIMS OF THE RESEARCH PROPOSAL

In this section, we review and summarize the original specific aims of the project, titled “Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient

Air Pollution,” which was awarded by HEI to a team at the Harvard T.H. Chan School of Public Health, with Francesca Dominici as the Principal Investigator (PI) and Antonella Zanobetti as the co-PI.

In late 2014, HEI issued a call for proposals (RFA 14-3) seeking studies to assess the health effects of long-term exposure to low levels of ambient air pollution with particular attention to (a) having sufficient size and statistical power to detect associations if they exist, (b) having the ability to test various potential confounders of these associations, and (c) exploring 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 few decades in North America, Europe, and other developed regions. Nonetheless, epidemiological studies continue to report associations of adverse health effects with air pollution even at these lower levels, and some recent studies have found associations at levels below current ambient air quality standards (Crouse et al. 2012; Hales et al. 2012; Shi et al. 2016). To inform future risk assessment and regulation, HEI committed funding (a) to confirm whether associations with adverse health effects continue to be observed as levels of air pollution decline further still and (b) to examine the shape of the ER function at those low levels; both are currently major uncertainties in air quality standards decisions.

As air pollution levels continue to decrease and regulatory actions become more costly, the quantification of the public health benefits of cleaner air become subject to increasing levels of scrutiny. Epidemiological analyses of claims data have provided strong evidence of air pollution 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 prior to our study, significant gaps in knowledge existed, particularly about the health effects of long-term exposure to lower levels of air pollution. First, there were not many studies before ours that investigated the health effects of long-term air pollution in areas with sparse monitoring (Aim 1). Second, the estimation of health effects associated with long-term exposure to low levels of air pollution presents key methodological challenges, including the fact (1) that 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) that health effects estimation at low exposure levels might be affected by a different set of confounders than at high exposure levels, (3) that information on individual-level potential confounders is limited in administrative data, (4) that estimation of the ER must account for potentially larger exposure errors at lower levels, (5) that identification of effect modifiers is challenged by the large number of possibilities that cannot all be tested individually, and (6) that 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, was necessary to investigate the health risks associated with exposure to low pollution levels and inform regulatory policy (Aim 2). Third, little was known about health effects at low pollution levels, not only mortality and morbidity outcomes, but also disease progression, particularly in highly susceptible populations, including children, the elderly, the disabled, pregnant women, and low-income adults (Aim 3). Fourth, methods for data sharing and reproducibility in air pollution epidemiology are of paramount importance, yet the scientific community has lacked tools to make them possible (Aim 4). To overcome these challenges, our team structured our work around four specific aims:

1. **Aim 1. Exposure prediction and data linkage** — 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 exposures to low levels of ambient $PM_{2.5}$ mass and components as well as the gaseous air pollutants O_3 and NO_2 , at high spatial resolution ($1\text{ km} \times 1\text{ km}$) for the continental United States during the period 2000–2016. Link these predictions to health data accounting for the misaligned nature of the data.
2. **Aim 2. Causal inference methods for estimating ER** — To estimate the whole ER by developing a new framework in Bayesian causal inference that is robust to model misspecification for confounding and to account for exposure error. Specifically, 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 the testing of each one individually.
3. **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 beneficiaries (28.6 million per year, 2000–2016); Medicaid beneficiaries (28 million per year, including 12 million children and 7 million disabled people, 2010–2014); and Medicare Current Beneficiary Survey (MCBS) beneficiaries (a nationally representative sample of approximately 15,000 beneficiaries per year with rich information on individual-level risk factors, including smoking, linked to Medicare claims). Examine the following health outcomes: (1) time to hospitalization by cause, (2) disease progression (time to rehospitalization), and (3) time to death.
4. **Aim 4. Tools for data access and reproducibility** — Develop tools for reproducible research, including approaches for data sharing, record linkage, and statistical software.

1.3. EXECUTIVE SUMMARY OF THE PHASE 1 REPORT

In the Phase 1 report (Dominici et al. 2019), we provided an overview of our work, grouped into the following five areas: (1) exposure prediction, (2) epidemiological studies of ambient exposures to air pollution at low levels, (3) sensitivity analysis, (4) methodological contributions in causal inference, and (5) open access research data platform. More specifically, we reported the following contributions:

1. **Exposure prediction for $PM_{2.5}$ (Aim 1)** — We summarized the exposure prediction modeling (Di et al. 2019) for estimating daily $PM_{2.5}$ levels at a resolution of $1\text{ km} \times 1\text{ km}$ for the continental United State for the period 2000–2012.
2. **Development of new statistical methods for causal inference (Aim 2)** — We summarized two sets of new methods for causal inference.
 - a. The first set consisted of new methods for using causal inference to account for exposure error, published by Wu and colleagues (2019) in *The Annals of Applied Statistics*. The code for the implementation of these methods is available at <https://github.com/wxwx1993/RC-GPS>. More specifically, in this work, we proposed a new approach for estimating causal effects when the exposure is imprecisely measured error and confounding adjustment is performed via a generalized propensity score (GPS). Using validation data, we proposed a regression calibration (RC)–based adjustment for a continuous error-prone exposure combined with GPS to adjust for confounding (RC–GPS). The outcome analysis is conducted after transforming the corrected continuous exposure into a categorical exposure. We considered confounding adjustment in the context of GPS subclassification, inverse probability treatment weighting, and matching. In simulations with varying degrees of exposure error and confounding bias, we reported that RC–GPS eliminates bias from exposure error and confounding compared with standard approaches that rely on the error-prone exposure. We applied RC–GPS to a rich data platform to estimate the causal effects of long-term exposure to $PM_{2.5}$ on mortality in New England for the period 2000–2012. The main study consisted of 2,202 zip codes covered by 217,660 $1\text{-km} \times 1\text{-km}$ grid cells with yearly mortality rates, yearly $PM_{2.5}$ averages estimated from a spatiotemporal model (error-prone exposure), and several potential confounders. The internal validation study included a subset of 83

- 1-km \times 1-km grid cells within 75 zip codes from the main study with error-free yearly $\text{PM}_{2.5}$ exposures obtained from monitor stations. Under assumptions of noninterference and weak unconfoundedness, and using matching, we found that exposure to moderate levels of $\text{PM}_{2.5}$ (8–10 mg/m^3 $\text{PM}_{2.5}$) causes a 2.8% (95% CI: 0.6%, 3.6%) increase in all-cause mortality compared with exposure to low levels ($\text{PM}_{2.5} < 8 \text{ mg}/\text{m}^3$).
- b. The second set consisted of new methods for using causal inference to flexibly estimate an ER function with local adjustment for confounding, published in a paper by Papadogeorgou and Dominici (2020) in *The Annals of Applied Statistics*. The R software package is available at <https://github.com/gpapadog/LERCA>. More specifically, in this work, we developed a Bayesian framework for the estimation of a causal ERC called LERCA (Local Exposure Response Confounding Adjustment). LERCA allows for (1) various confounders and various strengths of confounding at various exposure levels and (2) model uncertainty about confounder selection and the shape of the ER. LERCA also provides a principled way of assessing the observed covariates' confounding importance at various exposure levels. We compared our proposed method with state-of-the-art approaches in causal inference for ER estimation using simulation studies. We also applied the proposed method to a large data set for the entire United States that included health, weather, demographic, and pollution data for 5,362 zip codes for the years 2011–2013.
3. **Epidemiological studies (Aim 3)** — In the Phase 1 report, we summarized the following epidemiological studies.
 - a. **Short-term exposure to $\text{PM}_{2.5}$ and O_3 and all-cause mortality for the period 2000–2012.** The paper (Di et al. 2017a) was published in *The Journal of the American Medical Association*. Here, we conducted a case-crossover study to examine all deaths of Medicare beneficiaries in the continental United States from 2000 through 2012 and to estimate the mortality risk associated with short-term exposures to $\text{PM}_{2.5}$ and O_3 in the general population as well as in subgroups. The study was designed to estimate the association between daily mortality and air pollution at levels below the current daily U.S. National Ambient Air Quality Standards (NAAQS) to evaluate the adequacy of the current air quality standards for $\text{PM}_{2.5}$ and O_3 . We found that a 10- $\mu\text{g}/\text{m}^3$ daily increase in $\text{PM}_{2.5}$ and a 10-ppb daily increase in warm-season O_3 exposures were associated with a statistically significant increase of 1.42 and 0.66 deaths per 1 million per day, respectively. The risk of mortality remained statistically significant when restricting the analysis to days with $\text{PM}_{2.5}$ and O_3 levels much lower than the current U.S. daily NAAQS. The study included individuals living in smaller cities, towns, and rural areas that were unmonitored and thus had been excluded from previous time-series studies. There were no significant differences in the mortality risk associated with air pollution among individuals living in urban versus rural areas. These results provided evidence that short-term exposures to $\text{PM}_{2.5}$ and O_3 , even at levels much lower than the current daily standards, are associated with increased mortality, particularly for susceptible populations.
 - b. **Long-term exposure to $\text{PM}_{2.5}$ and O_3 and all-cause mortality for the period 2000–2012.** The paper (Di et al. 2017b) was published in *The New England Journal of Medicine*. Here, we conducted a nationwide cohort study of all Medicare beneficiaries from 2000 to 2012, a population of 61 million with 460 million person-years of follow-up. We used survival analysis (Andersen-Gill model) (Andersen and Gill 1982) to estimate the risk of death from any cause associated with long-term exposure (yearly average) to $\text{PM}_{2.5}$ concentrations lower than the current annual NAAQS (12 $\mu\text{g}/\text{m}^3$) and O_3 concentrations below 50 ppb. Subgroup analyses were conducted to identify populations with higher or lower pollution-associated risk of death from any cause. We found statistically significant evidence of adverse effects of $\text{PM}_{2.5}$ and O_3 exposures at concentrations below current national standards. This effect was greater for self-identified racial minorities and people with low income. Furthermore, we found that Black and Medicaid-eligible individuals had a much larger risk of death associated with exposure to $\text{PM}_{2.5}$ and O_3 than other subgroups (Di et al. 2017b). Medicare claims do not include individual-level data on behavioral risk factors, such as smoking and income, which could affect mortality and thus be important confounders. However, our analysis of the MCBS subsample did not find evidence of an association between smoking, income, and $\text{PM}_{2.5}$ or O_3 exposure. This important sensitivity analysis increased our level of confidence that lack of adjustment for these individual-level risk factors in the Medicare cohort did not lead to biased results. In another study, we analyzed a similar Medicare subsample with more detailed individual-level data on smoking, body mass index (BMI), and many other potential confounders linked to Medicare claims (Makar et al. 2017). In that analysis, we found that mortality and hospitalization risks of exposure to $\text{PM}_{2.5}$ were

not sensitive to the additional control of individual-level variables not available in the Medicare population as a whole.

1.4. CHALLENGES IDENTIFIED IN THE PHASE 1 REPORT

In the discussion of the Phase 1 report, we presented the strengths and limitations of our work. More specifically, we wrote, “[I]t is possible that our results could still be affected by unmeasured confounding bias, in particular, calendar time. In addition, although the methods that we have developed address several of the limitations of our current environmental epidemiological research paradigm, these new methods for causal inference are not easily scalable to massive data sets and may not be effective at dealing with continuous and time varying exposure.”

In the current report, we reanalyzed the data by Di and colleagues (2017b) using causal inference methods. This task required overcoming two enormous challenges: (1) developing new methods for causal inference in the context of estimating the causal effects of a continuous exposure and (2) applying these methods to a massive data set of 570 million observations for the period 2000–2016. Furthermore, we carefully accounted for the potential confounding of calendar year and addressed the sensitivity of our results to unmeasured confounding bias. Finally, we documented everything (sources of data, analytical data sets, and statistical code) to allow others to reproduce our results.

1.5. HARMONIZATION OF THE PHASE 1 AND FINAL REPORTS

In this section, we provide the itemized analyses (and analytic steps) conducted in the Phase 1 report but not in the final report, and vice versa. We also discuss why we decided to make some different analytical choices between the two reports. We do so separately for each of the key aims of the proposed project.

1. **Exposure assessment** — In the Phase 1 report, we summarized the exposure assessment approach for $PM_{2.5}$ (Di et al. 2019). In the final report, we also summarize the exposure assessment for O_3 and NO_2 (Di et al. 2020; Requia et al. 2020). All the modeling approaches were applied to data for the period 2000–2016. The estimated values for $PM_{2.5}$ are now publicly available at <https://beta.sedac.ciesin.columbia.edu/data/set/aqdh-pm2-5-concentrations-contiguous-us-1-km-2000-2016>.
2. **Methods for causal inference** — In the Phase 1 report, we summarized two methods for causal inference. The first introduced methods to account for exposure error, described above. This work is now being extended to analyses for the entire continental United States. Please note this is a very challenging task, and although the

analyses are underway, the results will not be available in time for the publication of this final report. Please see Section 5: Pipeline for Reproducible Research for a summary of this ongoing work. For the final report, we developed an alternative approach for the estimation of the causal ER function that is computationally tractable and scalable. This new approach is summarized in a preprint by Wu and colleagues (in review). We provide the details of this approach in Section 3 of the report.

- a. **Epidemiological studies** — In the Phase 1 report, we summarized epidemiological studies of the short- and long-term effects of $PM_{2.5}$ and O_3 exposure on all-cause mortality. As part of the work summarized in the final report (see below, in Sections 3, 4, and 5), we updated all the data to 2000–2016. We also summarized the long-term studies, and our analytical choices were different from those reported in the Phase 1 report, for the following reasons: (1) we wanted to reanalyze the entire national data set using established and new statistical methods for causal inference, (2) we wanted to address the comments of the Phase 1 report’s Low-Exposure Epidemiology Studies Review Panel as summarized in Section 1.4 above, (3) we needed to ensure computational scalability to handle more than 550 million observations, and (4) we needed to assess the sensitivity of the results to the analytical choices implemented in Phase 1 and to the new analytical choices for causal inference implemented in the final report. More specifically, we analyzed the entire national Medicare dataset, including Medicare beneficiaries for the period 2000–2016, using five statistical methods: (1) the same survival model (Andersen and Gill 1982) used by Di and colleagues (2017b) and summarized in the Phase 1 report, (2) a more computationally efficient Poisson formulation that is equivalent to the Andersen–Gill model under certain assumptions, and (3) three methods for causal inference based on the GPS. Two of these methods have been previously published, and one is a new method developed by our group (Wu et al. in review). Details are in Sections 3 and 4. In the final report, we have also summarized ongoing (unpublished) results for (1) long-term effects of $PM_{2.5}$ on mortality adjusted by O_3 and NO_2 , (2) long-term effects of $PM_{2.5}$ on mortality adjusted by NO_2 , (3) long-term effects of O_3 on mortality adjusted by $PM_{2.5}$ and NO_2 ; and (4) long-term effects of NO_2 on mortality adjusted by $PM_{2.5}$ and O_3 .
3. **Additional details on our analytical choices** — In the final report, and specifically for the long-term effects studies, we made the following analytical choices:
 - a. We analyzed the data for the period 2000–2016 (compared with 2000–2012 in Phase 1).

- b. For any statistical analysis, a unit of analysis must be defined. In our study, the unit of analysis could be either individuals or counts of individuals by zip code in a given year. For the Poisson and causal inference approaches in our study, also previously published (Wu et al. 2020), we decided to proceed with counts of individuals at the zip code-level in a given year as the unit of analysis, for the following reasons:
 - i. We have shown that the Anderson-Gill parameterization of the Cox model at the individual level is equivalent under certain assumptions to a Poisson model for counts of individuals at the zip code level in a given year. We included this in Section 3.3.
 - ii. The larger dataset created computational burdens in terms of data storage space and computer power to run the Cox regression model that was implemented in Phase 1. Traditional regression and causal inference analyses at the zip code level are computationally much more efficient than the same analysis at the individual level.
 - iii. The exposure assignment must be at the zip code level because the residential addresses of individual Medicare beneficiaries are only available at this level.
 - iv. The great majority of our potential measured confounders are from the U.S. Census and are available at the zip code level.
- c. As mentioned above, we reanalyzed the findings reported by Di and colleagues (2017b) using five statistical approaches to estimate the effect of $PM_{2.5}$ exposure on mortality, accounting for potential measured and unmeasured confounders. Details on the statistical approaches are provided in Section 3.
- d. We applied all five approaches to the data from 2000–2012, as was done by Di and colleagues (2017b), and to the data from 2000–2016 to assess the degree to which the results changed with more up-to-date data.
- e. To evaluate the model sensitivity to potential unmeasured confounders that vary over time, all five approaches were fitted twice, once with the year as a covariate (the main analysis) and once without (as a sensitivity analysis).
- f. To estimate low-level $PM_{2.5}$ effects on mortality, we applied the five statistical approaches, restricting analyses to the subpopulation of Medicare beneficiaries who were always exposed to $PM_{2.5}$ levels lower than $12 \mu\text{g}/\text{m}^3$ over the entire study period.
- g. The causal inference framework lends itself to the evaluation of covariate balance for measured confounders. The covariate balance indicates the quality of the causal inference approach at simulating randomized experiments and informs the degree to which one can make a valid causal assessment. Covariate balance was evaluated using mean absolute correlation (AC), with values <0.1 indicating high quality in simulating randomized experiments.
- h. We conducted further sensitivity analyses to unmeasured confounding by calculating the *E*-value. The *E*-value for a point estimate of interest (in our case, the HR) can be defined as the minimal strength of an association, on the risk ratio scale, that an unmeasured confounder would need to have with both the exposure and outcome, conditional on the covariates already included in the model, to fully explain the observed association under the null (Haneuse et al. 2019).
 - i. We also applied the newly developed method for the estimation of the causal exposure function (Wu et al. 2019) to estimate ERCs for $PM_{2.5}$, O_3 , and NO_2 adjusted by the other two pollutants (Section 4).
 - j. In Phase 1 of the project, we analyzed the MCBS data to assess the sensitivity of the results to omission of several individual-level confounders. Considering that the analyses (Makar et al. 2017) reported that estimation of the long-term effects of $PM_{2.5}$ on cause-specific hospitalization and all-cause mortality were not sensitive to the omission of several individual-level confounders (available in the MCBS but not in Medicare), we decided not to pursue these analyses further.

1.6. ROADMAP TO THE CONTENT OF THIS FINAL REPORT

Here we present an overall roadmap to the entire contents of the final report. In Section 2, we summarize the approaches for exposure assessment for $PM_{2.5}$, NO_2 , and O_3 for the period 2000–2016, which were published in the following papers.

1. Di Q et al. 2019. An ensemble-based model of $PM_{2.5}$ concentration across the contiguous United States with high spatiotemporal resolution. *Environ Int* 130:104909.
2. Di Q et al. 2020. Assessing NO_2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. *Environ Sci Technol* 54:1372–1384.
3. Requia WJ et al. 2020. An ensemble learning approach for estimating high spatiotemporal resolution of ground level ozone in the contiguous United States. *Environ Sci Technol* 54:11037–11047.

In Section 3, we summarize the national causal inference analysis of long-term effects of $PM_{2.5}$ on all-cause mortality among Medicare beneficiaries. We present methods and results. The methodological contribution and epidemiological studies are summarized in the following papers.

1. Wu X et al. 2020. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Sci Adv* 6:eaba5692.
2. Wu X et al. In review. Matching on generalized propensity scores with continuous exposures. <https://arxiv.org/pdf/1812.06575.pdf>.

In Section 4, we describe the application of methods for causal inference to estimate the ERCs to O_3 and NO_2 , and we provide results for the estimated ERCs for $PM_{2.5}$, O_3 , and NO_2 adjusted and not adjusted by the other two pollutants. This work has not been published yet.

In Section 5, we summarize the pipeline for reproducible research, including data and software. In Section 6, we summarize ongoing work. In Section 7, we provide a discussion of the strengths and limitations of the project.

2. EXPOSURE ASSESSMENT FOR $PM_{2.5}$, O_3 , AND NO_2

In this section, we summarize the exposure assessment approach for $PM_{2.5}$ (Di et al. 2019), NO_2 (Di et al. 2020), and O_3 (Requia et al. 2020) All of the modeling approaches were applied to data for the period 2000–2016 for the continental United States. The daily 1-km $PM_{2.5}$ predictions across the contiguous United States, 2000–2016, have been published on the National Aeronautics and Space Administration’s Socioeconomic Data and Applications Center website. The data are now publicly accessible and are available in both RDS and GeoTiff formats at <https://beta.sedac.ciesin.columbia.edu/data/set/aqdh-pm2-5-concentrations-contiguous-us-1-km-2000-2016>. The website is in beta test for the purpose of reviewing contents and finding bugs before final release. The beta test closed at the end of June 2021.

2.1. EXPOSURE ASSESSMENT FOR $PM_{2.5}$

In epidemiological analyses of air pollution health effects, the accurate estimation of $PM_{2.5}$ is an essential requirement. Different methods have been applied over the past decade to model $PM_{2.5}$, from typical linear regressions to machine learning approaches. We have developed and implemented an ensemble model that uses multiple machine learning algorithms to estimate daily $PM_{2.5}$ predictions at a 1-km \times 1-km grid resolution for the contiguous United States from January 1, 2000, to December 31, 2016.

The predictor variables that we included in the models were (1) $PM_{2.5}$ from monitoring stations obtained from the

U.S. EPA Air Quality System; (2) satellite-derived aerosol optical depth data; (3) 16 meteorological variables retrieved from the National Oceanic and Atmospheric Administration’s North American Regional Reanalysis data set; (4) land-use variables such as land-use coverage types, road density, restaurant density, elevation, and the normalized difference vegetation index to capture the impact of local emissions and air pollution levels; (5) several reanalysis datasets; and (6) daily predictions of total $PM_{2.5}$ mass and mass concentration of several $PM_{2.5}$ components from GEOS Chem, a global chemical transport model.

The modeling framework followed two steps, with each step combining a neural network, random forest, gradient boosting, and generalized additive model into an ensemble model. First, we modeled $PM_{2.5}$ separately with each algorithm on all input variables. The parameters of each algorithm were selected by cross-validated grid search processes. We blended the predicted concentration from each learner in a geographically weighted generalized additive model as an ensemble model to obtain $PM_{2.5}$ predictions. Some of the predictor variables had missing values; to obtain all input variables for the entire study area and during the entire study period, we imputed the missing values. We predicted the variables with missing values using the variables without missing values as predictor variables of a random forest. For some land-use variables, we filled missing values using linear interpolation because those variables were intermittent and unavailable over a certain period. In the final step, we applied the same three machine learning algorithms, but we included the spatially and temporally lagged $PM_{2.5}$ predictions from nearby monitoring sites and neighboring days together with existing predictor variables to obtain $PM_{2.5}$ daily predictions from 2000 to 2016 at each 1-km \times 1-km grid cell in the contiguous United States.

We validated the ensemble model with 10-fold cross-validation by training the model with 90% of the data and predicting $PM_{2.5}$ with the remaining 10% of monitoring sites. We then aggregated the cross-validated $PM_{2.5}$ from the 10 splits and compared them with the corresponding $PM_{2.5}$ monitoring values in each site on each day to obtain the total R^2 . We also computed the temporal R^2 (regressing the difference between predicted and monitored $PM_{2.5}$ in a site at a specific time with the annual mean in the same site) (Kloog et al. 2011), spatial R^2 (comparing the annual mean between monitored and predicted values in each site), root mean square error, and other metrics for model performance. It is also worth highlighting that 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).

Figure 1 shows the maps of $PM_{2.5}$ for 2000 and 2016. We obtained good model performance with a cross-validated R^2 of 0.86 for daily $PM_{2.5}$ predictions and a cross-validated R^2 of

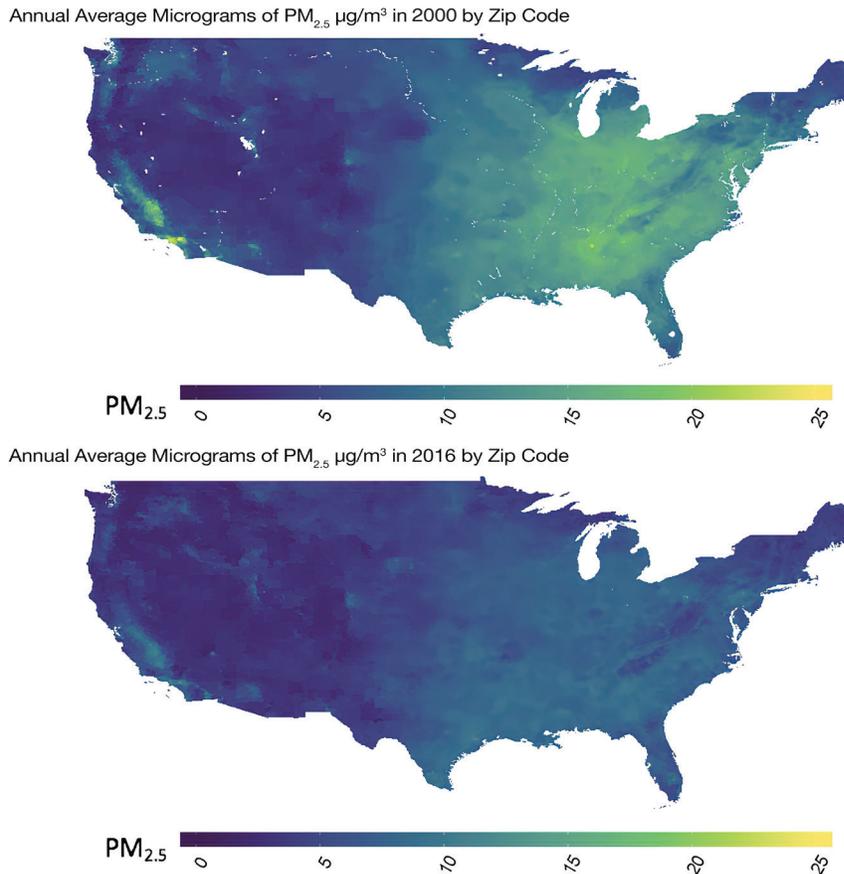


Figure 1. Annual average PM_{2.5} concentrations in the continental United States for 2000 and 2016. The white dots are due to missing data.

0.89 for annual PM_{2.5} estimates. The final model demonstrated good performance up to 60 mg/m³. The performance at lower levels was even better; most of the monitors had concentrations <12 mg/m³ (see next section for details). Therefore, when we calibrated the model to the monitors, the exposure error was lower at low concentrations where there were more data available to fit the models.

We found that the cross-validated R^2 of each machine learning algorithm varied by year, season, location, pollution concentration, and other factors, while the ensemble model, incorporating estimations from multiple machine learning algorithms, had a higher model performance. Therefore, applying a single method is not optimal for air pollution modeling.

We acknowledge that there is concern about the uncertainty of the exposure models, especially at low levels of exposure. Please note that most of the exposures and measurements of PM_{2.5} in the United States have been below the NAAQS for almost two decades. There is no reason to believe that modeled exposures have greater error in the exposure range where there are more data available to fit the models

than in the exposure range where the models rely on less data. As detailed above, we fitted our model to PM_{2.5} measurements at more than 1,900 monitoring stations. We compared the predicted annual average PM_{2.5} at each location with the measurements. The difference is the error in the estimated concentration. To see how the error variance changes with PM_{2.5} concentrations, we squared each error and then plotted the measured PM_{2.5} concentrations at those monitors in those years versus the smoothed squared errors. This approach estimates how the error variance changes with concentration. The results are shown in Figure 2. We found that the exposure error was smaller, not larger, at concentrations below or equal to 12 µg/m³. The PM_{2.5} exposures in earlier years are subject to greater uncertainty because the levels tended to be higher.

As detailed in Section 6, in which we summarize ongoing and future work, we are developing additional methods to incorporate exposure uncertainty into causal effect estimation. Furthermore, as part of our future work, we are planning to harmonize our analyses with the ones conducted by the Canadian and European teams also funded under this RFA (see Preface). As a part of the harmonization, we are planning

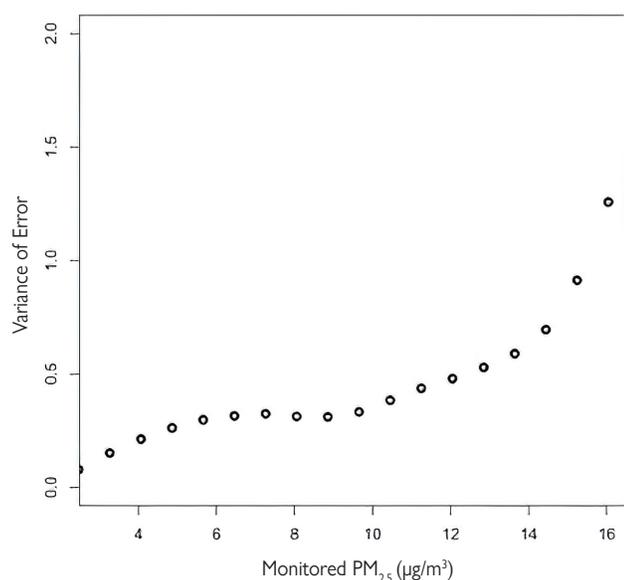


Figure 2. Variance of the exposure error plotted against $\text{PM}_{2.5}$ concentrations from monitors.

to rerun some of the key analyses using the $\text{PM}_{2.5}$ exposure estimates developed by the Canadian team, which relied on geographically weighted regression (Hammer et al. 2020; van Donkelaar et al. 2019).

2.2. EXPOSURE ASSESSMENT FOR NO_2

In addition to $\text{PM}_{2.5}$ exposure estimates, we also estimated daily NO_2 concentrations from 2000 to 2016 in a similar ensemble model-based approach. We fitted an ensemble model using a generalized additive model to combine estimates from three machine learning models — neural network, random forest, and gradient boosting — to obtain daily NO_2 predictions at 1-km-level grid cells, from 2000 to 2016 for the contiguous United States. Predictor variables included NO_2 column concentrations from satellite data, land-use variables, meteorological variables, predictions from two chemical transport models (GEOS-Chem and the regional-scale Community Multiscale Air Quality Model), and other ancillary variables.

Model training results for daily predictions from 2000 to 2016 indicated good model performance, with a 10-fold cross-validated R^2 of 0.788 overall, a spatial R^2 of 0.844, and a temporal R^2 of 0.729.

Figure 3 shows the maps of the predicted NO_2 concentrations for 2000 and 2016. We found a spatial clustering in the distribution of NO_2 , with the highest NO_2 concentrations in urban areas, especially in major cities, and along highways. We found that the model predicted well outside of major urban areas and also in rural areas. Compared with existing NO_2 models, our work shows that integrating many predictor variables and fitting algorithms can provide improved daily NO_2 predictions at high

spatial resolution, which should be useful in epidemiological studies of both long- and short-term exposures.

2.3. EXPOSURE ASSESSMENT FOR O_3

We estimated daily predictions for the daily maximum 8-hour O_3 at 1-km \times 1-km grid cells across the contiguous United States for the years 2000–2016, following the same model training process as described for $\text{PM}_{2.5}$ and NO_2 . Specifically, the models included 169 predictor variables, such as O_3 ground measurements from the EPA’s Air Quality System monitoring data, land-use variables, meteorological variables, chemical transport models and remote sensing data, and other data sources. After imputing missing data with machine learning algorithms, we applied a geographically weighted ensemble model that incorporated our three types of machine learners (neural network, random forest, and gradient boosting).

Similarly to the other pollutants, the monthly model uncertainty was estimated based on the difference between model predictions and observations, which considers exposure measurement error.

Figure 4 shows the maps of O_3 for 2000 and 2016. We obtained a 10-fold cross-validation R^2 of 0.902, a spatial R^2 of 0.862, and a temporal R^2 of 0.916, indicating good model performance. We found that the model performance of each machine learning algorithm was similar, while the ensemble model again had a higher performance. We also found that the model performed better in the East North Central region (R^2 of 0.93) and during summer (R^2 of 0.88) compared with the other regions and seasons. These predictions at a high temporal and spatial resolution should allow future studies to better estimate the health impacts of O_3 .

3. NATIONAL CAUSAL INFERENCE ANALYSIS ON LONG-TERM EFFECTS OF $\text{PM}_{2.5}$ ON ALL-CAUSE MORTALITY

Before we detail the methods and results of our national analysis of the long-term effects of $\text{PM}_{2.5}$ on all-cause mortality using causal inference methods, we introduce here the key ideas behind causal inference and explain their features. For more detail, please see Carone and colleagues (2020), Dominici and Zigler (2017), and Zigler and Dominici (2014). In addition, see the book by Imbens and Rubin (2015) and a recent paper by Dominici and colleagues (2020) for an overview of methods for causal inference.

3.1. WHY DO WE NEED CAUSAL INFERENCE METHODS IN ADDITION TO STANDARD REGRESSION APPROACHES?

Some scientists, including the former chair of the EPA’s Clean Air Scientific Advisory Committee, have argued against including studies that use traditional statistical approaches

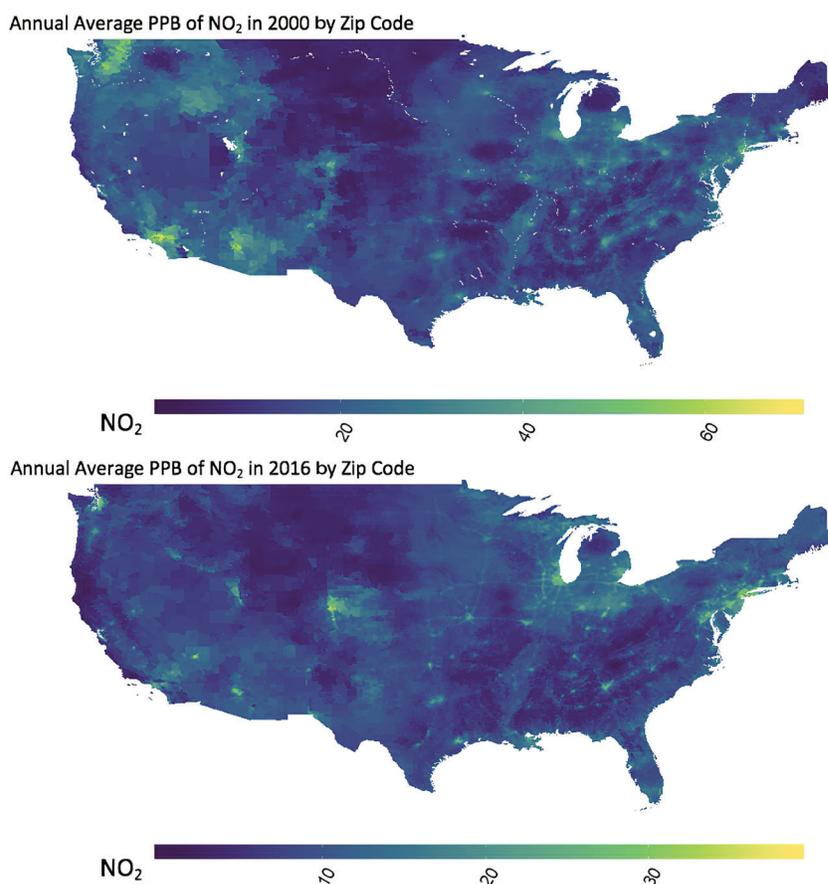


Figure 3. Annual average NO₂ concentrations in the continental United States for 2000 and 2016. The white dots are due to missing data.

to inform revisions of the NAAQS and propose focusing only on studies that use causal inference approaches (Environmental Protection Agency 2019). Their main criticism is that traditional approaches that include potential confounders as covariates in the regression model do not inform causality (Goldman and Dominici 2019).

Dominici and Zigler (Dominici and Zigler 2017; Zigler and Dominici 2014) previously discussed three notions of what constitutes evidence of causality in air pollution epidemiology. The first is causality inferred from evidence of biological plausibility. The second is consistency of results across many epidemiological studies and adherence to Bradford Hill causal criteria. The third is the use of causal inference methods that are more robust to model misspecification compared with traditional approaches and that, when assumptions are met, can isolate causal relationships. We believe that all three of these notions are valuable and essential. Indeed, we have stressed that the more comprehensive analysis of the full range of toxicological and epidemiological evidence serves as an important part of characterizing the complete picture of causality. Our work for this grant is mainly focused on

advancement of the causal inference methods described in the third notion of causality.

Causal inference methods have advantages and disadvantages compared with traditional regression methods. Their strengths are as follows:

1. Causal inference methods separate the design stage from the outcome analysis, thus increasing the objectiveness of causal analysis, and mimic a randomized experiment under a set of explicit identification assumptions.
2. They guide researchers to state explicitly all the identification assumptions needed for statistical analysis and equip them with a body of sensitivity analysis tools to understand how likely the identification assumptions are held (e.g., covariate balance and *E*-value).
3. They are more robust to model misspecification than traditional regression approaches.

But they also have limitations:

1. Causal inference methods often require increased computational resources due to the complexity of their algorithms.

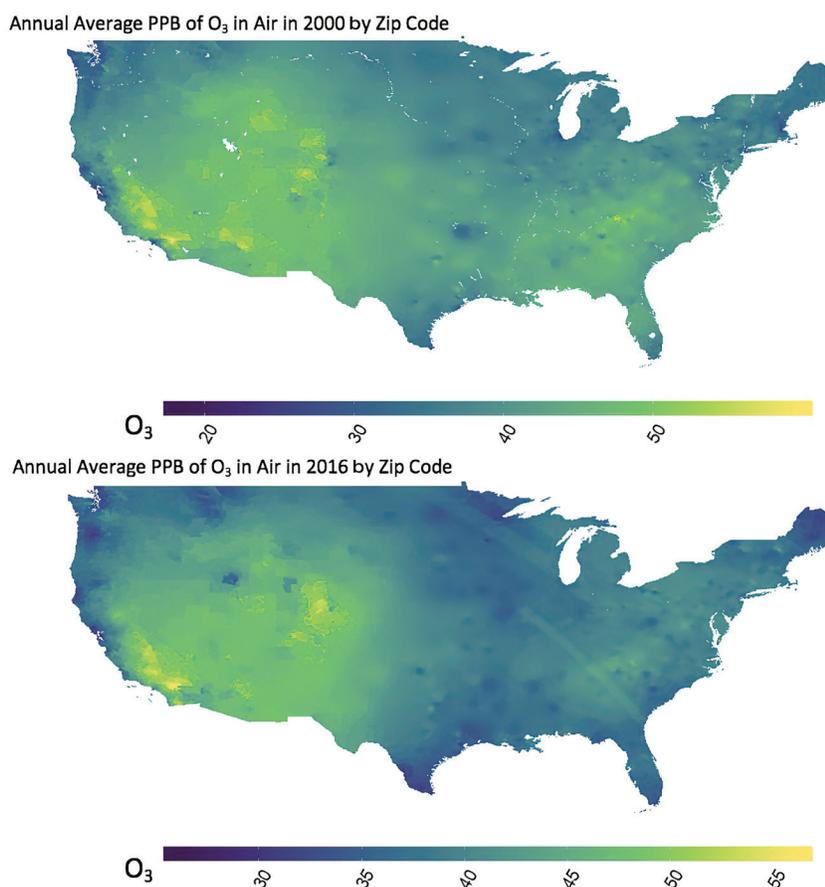


Figure 4. Annual average O_3 concentrations in the continental United States for 2000 and 2016. The white dots are due to missing data.

2. Some causal inference methods require steeper learning curves for new researchers because of the logical complexity and are often less familiar to many researchers.
3. Methods based on GPSs are still affected by unmeasured confounding bias.
4. Propagation of exposure error in health effects analyses under a causal inference framework are very challenging because error in the exposure also affects the propensity score. See Wu and colleagues (2019) for a broad description of the challenges and a proposed solution.

In a commentary led by Carone (Carone et al. 2020), we wrote, “We share the excitement of others that the discipline of causal inference has the potential to advance air pollution policy and allow the integration of modern statistical tools into air pollution epidemiology, but we also caution against unrealistic expectations by highlighting important difficulties ahead.” In a more recent paper (Dominici et al. 2020), which provided a tutorial on methods for causal inference, we wrote, “[W]e emphasize that experimental thinking is crucial in causal inference. The quality of the data (not necessarily the quantity), the study design, the degree to which the

assumptions are met, and the rigor of the statistical analysis allow us to credibly infer causal effects.”

In the current report, we pursued a causal inference approach because, under such an approach, we could quantify and visualize how closely we were able to approximate a randomized study. This can be accomplished by visualizing whether the measured confounders are balanced across exposed and nonexposed groups (Austin 2019; Imai and Ratkovic 2014) and how sensitive the results are to unmeasured confounding bias (Rosenbaum 2002). Loosely speaking, this framework allows us to assess how confidently we can make statements about causality using observational data under a set of explicit assumptions necessary for causal inference.

Under a causal inference framework, we articulate our research question using a potential outcome framework — that is, philosophically, we state a hypothetical causal question explicitly by mathematical formulas; for example, “If the pollution level is reduced from 12 units to 10 units, how many premature deaths can be saved?” We then borrow concepts from “randomized studies.” For example, under the ideal scenario of a clinical trial in drug development,

the investigator randomizes who receives the treatment and who does not and then evaluates whether there is a difference in measured outcomes between the two patient groups. This is called the causal effect because the differences in outcomes between the two groups are caused by the treatment. All other covariates are balanced because of the randomization. With the advanced methods for causal inference implemented in this report, we attempted to mimic this kind of randomized study using observational data (because randomized studies are unethical and unfeasible in the context of human subject environmental health studies). We could explicitly assess how closely we were able to approximate a randomized study and how sensitive our results were to unmeasured confounding bias. Thus, we can explain how our models tease out cause and effect.

Continuous Exposures Drastically Complicate Causal Inference Problems Most causal inference methods make the simplifying assumption of a binary exposure (or treatment). This assumption does not hold in environmental health research, because exposure to air pollution is a continuous variable. Approaches to estimating causal ERCs have been proposed, including methods that rely on the GPS (Wu et al. in review).

Innovation in Causal Inference As a methodological contribution to causal inference and in the context of a continuous exposure, we developed a robust and computationally efficient approach to nonparametrically estimate causal ERCs (Wu et al. in review). Next, we developed concepts and methods to understand covariate balance and to evaluate and overcome sensitivity to the critical assumptions of unmeasured confounding. The details of the approach are described below.

3.2 DATA

We obtained open cohort data for more than 68.5 million Medicare beneficiaries from 2000 to 2016 (Centers for Medicare & Medicaid Services), including demographic information on age, sex, race/ethnicity, date of death, and residential zip code. Each person was tracked through a unique patient ID.

3.2.1. Exposure Assessment

We estimated daily $PM_{2.5}$ levels at a high spatiotemporal resolution using a 1-km² grid network across the contiguous United States and a well-validated ensemble-based prediction model (Di et al. 2019). See Section 2 for details.

Residential addresses are not available for Medicare beneficiaries, only residential zip codes. For each standard zip code, we used zonal statistics to calculate the daily average $PM_{2.5}$ concentration based on all 1-km² grid cell predictions within the zip code via aggregations. More specifically, we first overlaid the zip code boundaries on the 1-km² grid cells and then averaged the predictions at 1-km² grid cells whose centroids fell within the boundary of that zip code (Zeiler 1999). For PO box-only zip codes, the average $PM_{2.5}$ concentrations

were calculated by linking to the predictions from the nearest 1-km² grid cell. Annual zip code averages were estimated by averaging the daily concentrations. We assigned the annual estimated zip code average $PM_{2.5}$ concentrations to individuals who lived in that zip code for each calendar year.

3.2.2. Potential Confounders

To adjust for confounding, we considered 10 zip code- and county-level confounders, including zip code-level socioeconomic status (SES) indicators from the 2000 and 2010 Census and the 2005–2012 American Community Surveys, and county-level information from the Centers for Disease Control and Prevention’s Behavioral Risk Factor Surveillance System (BRFSS). Specifically, we included (1) two county-level variables: average BMI and smoking rate; (2) eight zip code-level census variables: proportion of Hispanic residents, proportion of Black residents, median household income, median home value, proportion of residents in poverty, proportion of residents with a high school diploma, population density, and proportion of residents that own their house; and (3) four zip code-level meteorological variables: the summer (June–September) and winter (December–February) averages of maximum daily temperatures and relative humidity. We obtained the zip code-level meteorological variables using area-weighted aggregations based on daily temperature and humidity data on 4-km² gridded rasters from Gridmet via Google Earth Engine (Abatzoglou 2013; Gorelick et al. 2017). We also considered two variables indicating (1) the four Census geographic regions of the United States (Northeast, South, Midwest, and West) and (2) calendar years (2000–2016) to adjust for some residual or unmeasured spatial and temporal confounding, respectively.

3.2.3. Data Linkage

Outcome data were available at the postal zip code level, at which we also assigned annual $PM_{2.5}$ exposures. Outcome and exposure information were available for 35,924 of the 40,431 zip codes in the 48 contiguous states and Washington, DC. We then mapped potential confounders at zip code tabulation areas to postal zip codes to link the outcome and exposure data to potential confounders obtained from the U.S. Census, American Community Surveys, and the BFRSS. The total number of zip codes included in our main analysis with information on all outcome, exposure, and confounder data was 31,337.

In Section 6, titled “Pipeline for Reproducible Research,” we provide detail on the variables included in the analyses and the R code for the implementation of the methods. All study data sources are publicly available.

3.3. METHODS

3.3.1. Study Population

Table 1 shows the characteristics of the study cohort. Our study population consisted of more than 68.5 million

Table 1. Characteristics of the Study Cohorts^a

Variables	Entire Medicare Enrollees	Medicare Enrollees Exposed to PM _{2.5} ≤ 12 µg/m ³
Number of individuals	68,503,979	38,366,800
Number of deaths	27,106,639	10,124,409
Total person-years	573,370,257	259,469,768
Median years of follow-up	8.0	8.0
Individual-level characteristics		
Age at entry (years)		
65–74 (%)	80.6	88.1
75–84 (%)	14.9	9.0
85–94 (%)	4.1	2.6
95 or above (%)	0.4	0.2
Mean (SD)	69.2 (6.7)	67.6 (5.6)
Sex		
Female (%)	55.5	53.8
Male (%)	44.5	46.2
Race		
White (%)	83.9	84.7
Black (%)	9.1	7.3
Asian (%)	1.8	1.8
Hispanic (%)	2.0	2.2
North American Native (%)	0.3	0.4
Medicaid eligibility (%)	11.7	10.9
Area-level risk factor characteristics		
Ever smoked (%)	47.3	47.3
Below poverty level (%)	10.5	10.1
Below high school education (%)	28.5	25.6
Owner-occupied housing (%)	72.0	72.9
Hispanic (%)	8.9	7.5
Black (%)	8.9	9.2
Population density (persons/km ²)	600.0	489.1
Mean BMI (kg/m ²)	27.6 (1.1)	27.6 (1.1)
Median household income (\$1,000)	48.9 (21.7)	50.3 (22.0)
Median home value (\$1,000)	162.5 (140.9)	170.9 (146.2)
Meteorological variables		
Summer temperature (°C)	29.5 (3.7)	29.5 (3.9)
Winter temperature (°C)	7.6 (7.2)	7.4 (7.6)
Summer relative humidity (%)	88.0 (11.7)	86.7 (12.7)
Winter relative humidity (%)	86.2 (7.3)	86.4 (7.6)
PM_{2.5} concentrations (µg/m³)	9.8 (3.2)	8.4 (2.3)

^aMean (SD) is shown for continuous variables.

Medicare enrollees between 2000 and 2016. Of those, more than 38 million were exposed to $PM_{2.5}$ levels below or equal to $12 \mu\text{g}/\text{m}^3$. More than 27 million deaths occurred during the study period. 11.7% of the study population was also eligible for Medicaid. Medicare claims data, obtained from the Centers for Medicare & Medicaid Services (Centers for Medicare & Medicaid Services), are an open cohort, including demographic information such as age, sex, race/ethnicity, date of death, and residential zip code. A unique patient ID was assigned to each person to allow tracking over time. Medicare beneficiaries entered our cohort in 2000 if enrolled before 2000 or upon their enrollment after 2000. After enrollment, each beneficiary was followed annually until the year of their death or the end of our study period (31 December 2016). This study was conducted under a protocol approved by the Harvard T.H. Chan School of Public Health Human Subjects Committee.

3.3.2. Statistical Analysis

In this section, we provide mathematical details on our statistical analyses. The *R* code for the implementation of all five statistical approaches is published and available at <https://github.com/NSAPH/National-Causal-Analysis>.

We implemented five statistical approaches to estimate the effect of $PM_{2.5}$ exposure on mortality, accounting for potential confounders. The two traditional approaches rely on regression modeling for confounding adjustment: (1) the Cox proportional hazards model and (2) Poisson regression. We created multiple observations for each subject, each representing a person–year of follow-up. We fitted the Cox hazards models, using follow-up year as the time metric and annual $PM_{2.5}$ as the time-varying exposure, stratifying by age (5-year categories), sex, race/ethnicity, and Medicaid eligibility (a surrogate for individual-level SES). We fitted the Poisson models for aggregated outcomes (i.e., counts of death) by zip code and year, stratified by the same individual-level characteristics and follow-up year. We adjusted both for confounding by including 10 zip code– or county-level risk factors, four zip code–level meteorological variables, and indicators for geographic region (Northeast, South, Midwest, and West). To account for long-term time trends, we included calendar year as a categorical variable. The details of the models are provided below.

Cox Proportional Hazard Approach We fitted stratified Cox proportional hazards models using annual $PM_{2.5}$ as the time-varying exposure and stratifying by four individual-level characteristics. We included all individual-level variables for Medicare beneficiaries that are available in Medicare Part A data. We were able to account for important individual-level characteristics (age, race, sex, Medicaid eligibility) by fitting a stratified Cox proportional hazards model. In our main analysis, we adjusted for 14 zip code– or county-level time-varying covariates as well as a dummy region variable

and a dummy calendar year variable. The Cox proportional hazards survival model was specified as: Survival (follow-up year, death) $\sim PM_{2.5}$ + area-level risk factors + meteorological variables + dummy year + dummy region + strata (age, race, sex, Medicaid eligibility).

Poisson Regression Approach We fitted the Poisson regression model using annual $PM_{2.5}$ as the time-varying exposure; the count of deaths at the given follow-up year, calendar year, and zip code as the outcome; and the corresponding total person–time as the offset term. To adjust for potential confounding, we included the same 14 zip code– or county-level time-varying covariates, dummy region variable, and dummy calendar year variable as those included in the Cox proportional hazards models. We used a stratified Poisson regression model formulation to account for the strata-specific baseline risk rates by stratifying on individual-level characteristics. The Poisson regression model was specified as: $\log(E[\text{death counts}]) \sim PM_{2.5}$ + area-level risk factors + meteorological variables + dummy year + dummy region + strata (age, race, sex, Medicaid eligibility, follow-up year) + offset ($\log[\text{person-year}]$).

Equivalency Between Cox and Poisson The key difference between the Cox proportional hazards model and the Poisson model is that the Poisson model is fitted on an aggregated dataset while the Cox survival model is fitted on an individual-level dataset. However, because in the Cox model the common exposure and the set of potential confounders are assigned at zip code or county level, we can still construct the aggregated dataset stratified by individual-level characteristics, zip code, and follow-up time without losing the ability to examine individual-level health effects caused by common exposures using stratified Poisson models.

Below we show the mathematical details on the equivalence between the Cox proportional hazards model and the stratified Poisson model. Let z denote each zip code, a denote each follow-up year, and t denote each calendar year. $W_{z,t}$ denotes the annual average $PM_{2.5}$ concentration in zip code z in calendar year t . $C_{z,t}$ denotes the zip code time-varying covariates, which serve as potential confounders, in calendar year t and zip code z . We fitted the following stratified Cox proportional hazards model to examine the long-term effects of $PM_{2.5}$ on the health outcomes.

$$h^{c,z}(a,t) = h_0^c(a) \exp(\beta_1 W_{z,t} + \beta_2 C_{z,t}) \quad (1)$$

where $h^{c,z}(a,t)$ denotes the hazard of mortality at follow-up year a , calendar year t , and zip code z for individual-characteristic strata c (i.e., age group, sex, race, Medicaid eligibility), and $h_0^c(a)$ is a strata-specific baseline hazard function.

Model (1) can be written as

$$\frac{E(Y_{a,t}^{c,z})}{T_{a,t}^{c,z}} = h_0^c(a) \exp(\beta_1 W_{z,t} + \beta_2 C_{z,t}) \quad (2)$$

where $E(Y_{a,t}^{c,z})$ denotes the expected number of deaths at follow-up year a , calendar year t , and zip code z for each individual-characteristic stratum c , and $T_{a,t}^{c,z}$ is the corresponding total person-time in that stratum.

Taking the log of both sides, model (2) can be written as

$$\log(E(Y_{a,t}^{c,z})) = \log(T_{a,t}^{c,z}) + \log(h_0^c(a)) + \beta_1 W_{z,t} + \beta_2 C_{z,t} \quad (3)$$

The Cox proportional hazards model (1) is equivalent to the stratified Poisson model (3). We also considered three approaches for causal inference that rely on the potential outcome framework and GPS. These approaches adjust for confounding using (1) matching by GPS, (2) weighting by GPS, and (3) adjustment by GPS, by including GPS as a covariate in the health outcome model.

For any statistical analysis, a unit of analysis must be defined. For the Poisson and causal inference approaches, a study unit is defined as counts of individuals by zip code in a given year. We have shown that the Anderson-Gill parameterization of the Cox model at the individual level is equivalent under certain assumptions to a Poisson model for counts of individuals at the zip code-level in a given year. Traditional regression and causal inference analyses at the zip code level are computationally much more efficient than the same analysis at the individual level. The exposure assignment must be at the zip code level because the residential address of each Medicare beneficiary is only available at this level. The great majority of our potential measured confounders are from the U.S. Census and available at zip code level.

Below we summarize the methods for causal inference. The causal inference analysis is conducted in two stages: a design stage and an analysis stage. In the design stage, we create a pseudo-population that is meant to mimic a randomized control study.

Causal Assumptions The GPS approach relies on a potential outcome framework (Hirano and Imbens 2004). Briefly, a potential outcome is an outcome that would have been realized if an individual had received a specific value of the exposure. Using propensity scores (Turner et al. 2016) to adjust for confounding in a potential outcome framework is one very common approach for studying causal effects in observational studies.

GPS allows one to simultaneously balance a large set of covariates in the exposed and reference populations. By ensuring covariate balance between the exposed population and a reference population, a pseudo-population is created that mimics a randomized experiment. Randomized experiments are considered the gold standard to inform causality and ensure that the covariate distributions do not differ by exposure status (i.e., that the covariates are balanced).

Following the potential outcome framework, we state three assumptions of causal identification in the GPS approach.

The first assumption is consistency. This assumption is also referred to as no-interference or the stable-unit-treatment-value assumption. In brief, we assume that the potential outcome for a given observation is not affected by the exposure of any other unit and that each exposure defines a unique outcome for each observation.

In the current study, we assumed that air pollution concentrations would not affect the health outcome of an individual who lives in an adjacent zip code. We believe that the individual's health outcomes will be dominantly affected by exposure to air pollution in the zip code where they live. We cannot rule out effects from adjacent zip codes either. The question of a spillover effect is of scientific interest and needs further investigation.

The second assumption is overlap. This assumption, sometimes referred to as the positivity assumption, states that the exposure is not assigned deterministically and thus that each individual has a positive chance of receiving any exposure level, regardless of the individual's covariate profile (i.e., set of potential confounders).

In the current study, each zip code could hypothetically be exposed to any level of air pollutants. We believe this assumption is likely to hold theoretically. Indeed, in practice, the probability of a zip code with a very low air pollution concentration encountering a very high concentration could be small.

The third assumption is weak unconfoundedness. This assumption, sometimes referred to as the no-unmeasured-confounder assumption, states that the mean potential outcome under a specific exposure level is the same across every exposure level once one conditions on potential confounders (i.e., exposure assignment is unrelated to potential outcomes within strata created by potential confounders). This assumption indicates the possibility that, if sufficiently many relevant covariates that characterized the individual's profile are collected, we would be able to approximate a stratified randomized experiment from observational studies by conditioning on the set of covariates.

In the current study, we collected as many covariates associated with both the exposure and outcomes as we could to adjust for measured confounders. Also, we included both year and region as indicator variables to adjust for some unmeasured confounders that co-vary temporally or spatially with both the exposure and the outcome. We conducted a sensitivity analysis to unmeasured confounding by calculating the E -value. We believe the threat of unmeasured confounding to our analysis results was minimized, given our use of state-of-the-art statistical methods to control and assess the potential confounding. Yet, because this assumption was not directly verifiable based on the observed data, we cannot rule out the possibility that unmeasured confounding exists.

The main advantage of causal inference approaches compared with more traditional approaches is that their design

and analysis stages are separate (Rubin 2008; Stuart 2010). In the design stage, investigators design the study, creating a pseudo-population that mimics a randomized experiment, without using the outcome information. Only after the design stage is complete does the analysis stage begin, conducting outcome analysis on the pseudo-population. We considered the following three common and validated causal inference approaches: (1) GPS matching, (2) GPS weighting, and (3) GPS adjustment.

GPS Estimation The three proposed causal inference approaches required the *estimation* of GPS as their first step. In our study, we modeled the conditional density of exposure (i.e., zip code–level annual average $PM_{2.5}$) on the 14 zip code– or county-level time-varying covariates, as well as a dummy region variable and a dummy calendar year variable, by using a gradient boosting machine with normal residuals (Chen and Guestrin 2016; Zhu et al. 2015). The gradient boosting machine model is specified as: $PM_{2.5} \sim \text{area-level risk factors} + \text{meteorological variables} + \text{dummy year} + \text{dummy region} + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$.

GPS Matching The GPS matching approach was newly developed by our team and is described in detail by Wu and colleagues (in review). The ultimate objective for matching is to construct matched datasets that approximate a randomized experiment as closely as possible by achieving good covariate balance. In the continuous exposure setting, the challenge is that it is unlikely that two units will have the exact same level of exposure; thus, it is infeasible to create a finite sample representing a quasi-experimental arm with the same exposure level solely by matching on GPS. Therefore, we proposed a nearest-neighbor caliper-matching procedure with replacement, which jointly matches on both the estimated GPS and the exposure values. The closeness of exposure level guarantees that the matched unit is a valid representation of observations for a particular exposure level, while the closeness of GPS ensures that we were properly adjusting for confounding. Importantly we assessed covariate balance in the matched population, and if covariate balance was achieved, we fitted a univariate Poisson regression model regressing the death counts with an offset person–time term, on the exposure $PM_{2.5}$, stratifying by four individual-level characteristics and the same follow-up year. The Poisson regression model was specified as: $\log(E[\text{death counts}]) \sim PM_{2.5} + \text{strata}(\text{age, race, sex, Medicaid eligibility, follow-up year}) + \text{offset}(\log[\text{person year}])$, on the matched pseudo-population.

Mathematical Details of the GPS Matching Let N denote the study sample size. For each study unit $j = 1, 2, \dots, N$, W_j denotes the annual average $PM_{2.5}$ concentration for unit j . C_j denotes the zip code- or county-level time-varying covariates, which serve as potential confounders for unit j , including area-level risk factors, meteorological variables, year, and region. $Y_j(w)$ denotes the counterfactual outcome for unit j at the exposure level w (i.e., the anticipated death counts for

unit j had the annual average $PM_{2.5}$ been at level w). In our analysis, a study unit is defined as a zip code z in a year t . Wu and colleagues (in review) showed the mathematical details of our newly developed matching approach.

We proposed a cross-validation procedure to find the optimal caliper δ that minimizes the covariate balance in the resulting matched pseudo-population. The optimized caliper results in $L = 50$ levels of exposure in our study.

GPS Weighting Following Robins and colleagues (2000), our weighting approach involves using the inverse of the GPS to weigh the observations. To stabilize the weights, we multiply the inverse of the GPS by the marginal probabilities of exposure. We constructed the weighted pseudo-population by using the inverse of estimated GPS to weigh each observation. We first checked the covariate balance on the weighted pseudo-population, and if covariate balance was achieved, we fitted a weighted univariate Poisson regression model regressing the death count with offset term as the person–time on $PM_{2.5}$ exposure, incorporating the assigned weights and stratifying by the four individual-level characteristics and the same follow-up year. The Poisson regression model is specified as: $\log(E[\text{death counts}]) \sim PM_{2.5} + \text{strata}(\text{age, race, sex, Medicaid eligibility, follow-up year}) + \text{offset}(\log[\text{person year}])$, weights = $f(PM_{2.5})/GPS$, where $f(PM_{2.5})$ is the marginal density function of exposure $PM_{2.5}$, which serves as a stabilizing term (Robins et al. 2000).

GPS Adjustment Following Hirano and Imbens seminal paper (2004), our covariate adjustment approach included the estimated GPS as a covariate in the outcome model. Hirano and Imbens showed that including the estimated GPS as a covariate together with the exposure in a bivariate outcome model can remove confounding bias when estimating causal ERCs. We modeled the conditional expectation of the death counts given the exposure and the estimated GPS as a stratified Poisson regression with flexible formulation of bivariate variables, with the corresponding person–time offset. The outcome model was specified as: $\log(E[\text{death counts}]) \sim PM_{2.5} + PM_{2.5} \times GPS + GPS + GPS^2 + \text{strata}(\text{age, race, sex, Medicaid eligibility, follow-up year}) + \text{offset}(\log[\text{person year}])$. The form of the outcome model for GPS adjustment comes from Hirano and Imbens (2004), who proposed this outcome model. This is not the only or necessarily the best form of the outcome model but is one that has been proposed in the literature. In contrast to the GPS matching/weighting approaches, where the analysis is complete after fitting the Poisson regression model, for the GPS adjustment approach the coefficients from the Poisson regression model do not provide any causal interpretation; instead, the causal outcome analysis is conducted on the counterfactuals predicted by the Poisson model (Hirano and Imbens 2004). We fitted a univariate linear regression model regressing the counterfactual mean hazard rates for each $PM_{2.5}$ level, stratifying by four individual-level characteristics and the same follow-up year. The outcome

linear regression model is specified as: $E(\text{hazard rates}) \sim \text{PM}_{2.5} + \text{strata}(\text{age, race, sex, Medicaid eligibility, follow-up year})$.

All five approaches were fitted on the 2000–2016 data. The 2000–2016 cohort consisted of 68,503,979 subjects (573,370,257 person-years); there were 27,106,639 deaths (39.6%; Table 1). For all models, we performed a stratified outcome model analysis by four individual-level characteristics: (1) a 5-year category of age at entry (65 to 69, 70 to 74, 75 to 79, 80 to 84, 85 to 89, 90 to 94, 95 to 99, and above 100 years of age); (2) race/ethnicity (White, Black, Asian, Hispanic, Native American, and other); (3) sex (male or female); and (4) an indicator variable for Medicaid eligibility (a surrogate for individual-level SES).

Covariate Balance For the GPS-based approaches, we assessed the quality of our study design and, in particular, evaluated the covariate balance for the constructed pseudo-population via absolute correlation. Our balance diagnostics were motivated by the balancing property of the GPS. The key is that if two variables are independent of one another, then the correlation between them will be zero. Covariate balance was evaluated using mean AC, with values <0.1 indicating high quality in simulating randomized experiments (Austin 2019; Wu et al. in review).

Evaluation of Unmeasured Confounding We conducted a sensitivity analysis to evaluate the robustness of our results to unmeasured confounding by calculating the E -value (Haneuse et al. 2019). The E -value for the point estimate of interest (in our case, the HR) can be defined as the minimal strength of an association, on the risk ratio scale, that an unmeasured confounder would need to have with both the exposure and outcome, conditional on the covariates already included in the model, to fully explain the observed association under the null. We calculated the E -values for our reported HRs per $10\text{-}\mu\text{g}/\text{m}^3$ increase of long-term exposure to $\text{PM}_{2.5}$. The calculation of E -values can be implemented through the E -value calculator by Mathur and colleagues (2018), available at <https://www.evalue-calculator.com/>.

Uncertainty Quantification We evaluated the 95% CIs for all models by m -out-of- n subsampling block bootstrap to account for spatial correlation. We conducted the subsampling bootstrap with zip codes as the block units; that is, each bootstrap replicate contains a smaller number of zip codes compared with the original data set. Subsampling bootstrap was used to handle scenarios where the “i.i.d.” (independent and identically distributed) full bootstrap fails (Politis 2003). We used zip code as the sampling unit in the block bootstrap. We accounted for the correlation between observations within the same zip code by the “block” nature of the bootstrap procedure. By randomly sampling zip codes for each bootstrap replicate, we broke down spatial dependence given covariates. Therefore, it was less likely that our results were affected by spatial correlation. We recalculated the GPS and refitted the outcome model in each bootstrap replicate to

ensure that the bootstrapping procedure jointly accounted for the variability associated both with the GPS estimation and with the outcome model.

Total Events Avoided We estimated the total number of deaths that would have been avoided among the elderly per decade if all areas were in compliance with the previous World Health Organization guidelines ($\leq 10\text{ }\mu\text{g}/\text{m}^3$ annual $\text{PM}_{2.5}$ exposure) (World Health Organization 2005). Nethery and colleagues (2019) identified, named, and defined the causal quantity Total Events Avoided (TEA) as the difference in the expected number of health events under counterfactual pollution exposures and the observed number of health events under factual pollution exposures. Such a causal quantity is particularly related to the health policy that aims to answer the question “How many deaths were avoided in the Medicare population per decade due to NAAQS changes in $\text{PM}_{2.5}$ over the same time period?”

We created the counterfactual $\text{PM}_{2.5}$ exposures if all zip codes in the continental United States had complied with the previous World Health Organization guidelines ($\leq 10\text{ }\mu\text{g}/\text{m}^3$ annual $\text{PM}_{2.5}$). For zip codes that did not comply with the standard until 2016, their counterfactual was assumed to be exposure exactly at this hypothesized standard ($10\text{ }\mu\text{g}/\text{m}^3$). This was a conservative estimate, because it answered the question of TEA if these zip codes were exactly at $10\text{ }\mu\text{g}/\text{m}^3$ and not lower. For zip codes already in compliance, we assumed their concentration was unchanged, which otherwise would result in an even higher TEA.

We compared this counterfactual scenario with the factual scenarios during the most recent decade (2007–2016). For zip codes with annual $\text{PM}_{2.5}$ concentration $>12\text{ }\mu\text{g}/\text{m}^3$, the numbers for the TEA were obtained using the most conservative HR from our main analysis (HR = 1.068 [95% CI, 1.054 to 1.083]; see Table 2). For zip codes with annual $\text{PM}_{2.5}$ concentration $10\text{--}12\text{ }\mu\text{g}/\text{m}^3$, the numbers for the TEA were obtained using the most conservative HR from our low-level analysis (HR = 1.23 [95% CI, 1.18 to 1.28]; see Table 2). Zip codes with annual $\text{PM}_{2.5}$ concentration $<10\text{ }\mu\text{g}/\text{m}^3$ did not contribute to the TEA. For the CI calculation, we used the lower and upper bounds of the 95% CIs from the HR estimates (which were obtained by bootstrap).

3.4. RESULTS

Our causal inference framework lends itself to the evaluation of covariate balance for measured confounders. The covariate balance indicates the quality of the causal inference approach in simulating randomized experiments and informs the degree to which one can make a valid causal assessment. Covariate balance was evaluated using mean AC, with values <0.1 indicating high quality in simulating randomized experiments. Figure 5 shows that the AC was smaller than 0.1 using causal inference GPS methods (matching and weighting), thus strengthening the interpret-

Table 2. HRs and 95% CIs

Cohort	Methods	Main Analysis	Not Adjusted for Year	Not Adjusted for Meteorological Variables
2000–2016	Matching	1.068 (1.054, 1.083)	1.089 (1.075, 1.103)	1.077 (1.063, 1.092)
	Weighting	1.076 (1.065, 1.088)	1.144 (1.134, 1.154)	1.087 (1.076, 1.098)
	Adjustment	1.072 (1.061, 1.082)	1.115 (1.103, 1.128)	1.061 (1.050, 1.072)
	Cox	1.066 (1.058, 1.074)	1.172 (1.164, 1.180)	1.058 (1.050, 1.066)
	Poisson	1.062 (1.055, 1.069)	1.166 (1.158, 1.174)	1.057 (1.049, 1.064)
2000–2016 Low Level ^a	Matching	1.261 (1.233, 1.289)	1.318 (1.287, 1.349)	1.251 (1.222, 1.280)
	Weighting	1.268 (1.237, 1.300)	1.387 (1.355, 1.419)	1.262 (1.232, 1.291)
	Adjustment	1.231 (1.180, 1.284)	1.424 (1.327, 1.527)	1.233 (1.169, 1.299)
	Cox	1.369 (1.340, 1.399)	1.569 (1.536, 1.602)	1.358 (1.330, 1.387)
	Poisson	1.347 (1.320, 1.375)	1.541 (1.510, 1.573)	1.343 (1.316, 1.370)
2000–2012	Matching	1.055 (1.042, 1.068)	1.085 (1.072, 1.098)	
	Weighting	1.067 (1.056, 1.079)	1.114 (1.103, 1.125)	
	Adjustment	1.047 (1.037, 1.057)	1.078 (1.065, 1.090)	
	Cox	1.059 (1.051, 1.067)	1.128 (1.120, 1.136)	
	Poisson	1.055 (1.048, 1.063)	1.123 (1.116, 1.131)	
2000–2012 Low Level ^a	Matching	1.271 (1.241, 1.301)	1.293 (1.262, 1.324)	
	Weighting	1.298 (1.254, 1.344)	1.383 (1.343, 1.425)	
	Adjustment	1.233 (1.176, 1.292)	1.385 (1.291, 1.485)	
	Cox	1.367 (1.331, 1.404)	1.538 (1.497, 1.580)	
	Poisson	1.342 (1.308, 1.377)	1.509 (1.471, 1.548)	

^a Low level = Medicare beneficiaries exposed to $PM_{2.5} \leq 12 \mu\text{g}/\text{m}^3$. (From Wu et al. 2020, © the Authors, some rights reserved; exclusive licensee AAAS. Distributed under a CC BY-NC 4.0 License.)

ability and validity of our analyses as providing evidence of causality.

Figure 6 summarizes the effect estimates for the period 2000–2016. The effect estimates are presented as HRs per $10\text{-}\mu\text{g}/\text{m}^3$ increase in annual $PM_{2.5}$. 95% CIs for all models were evaluated by m-out-of-n block bootstrap to account for spatial correlation. More specifically, we recalculated the GPS and refitted the outcome model in each bootstrapped sample to ensure that the bootstrapping procedure jointly accounted for the variability associated both with the GPS estimation and the outcome model. For the period 2000–2016, we found that all statistical approaches provided consistent results: a $10\text{-}\mu\text{g}/\text{m}^3$ decrease in $PM_{2.5}$ led to a statistically significant decrease in mortality rate ranging between 6% and 7% ($= 1 - 1/\text{HR}$) (HR estimates 1.06 [95% CI, 1.05 to 1.08] to 1.08 [95% CI, 1.07 to 1.09]). The estimated HRs were larger when studying the cohort of Medicare beneficiaries that were always exposed to $PM_{2.5}$ levels lower than $12 \mu\text{g}/\text{m}^3$ (1.23 [95% CI, 1.18 to 1.28] to 1.37 [95% CI, 1.34 to 1.40]). Our results are consistent with recent epidemiological studies reporting a stronger association between long-term exposure to $PM_{2.5}$ and adverse health

outcomes at exposure levels below the national standards, suggesting no safe threshold for harmful pollution (Shi et al. 2016; Villeneuve et al. 2015). It is also important to point out that, when estimating HRs at levels below or equal to $12 \mu\text{g}/\text{m}^3$, the causal inference approaches produce smaller estimates of the HRs than the traditional regression. We hypothesize that this might be due to model misspecification of the traditional regression (which assumes that the confounding adjustment is linear), whereas in the context of the GPS, we do not need to make this assumption. Also, the low values of the AC in Figure 5 reassure us that the GPS approaches provided an adequate adjustment for measured confounding bias. Unfortunately, similar diagnostics for a regression model cannot be implemented.

We also found a statistically significant link between $PM_{2.5}$ exposures and all-cause mortality for the period 2000–2012 (Table 2), showing the consistency of the scientific evidence with Di and colleagues (2017b). The estimated HRs were obtained under four cohorts using five different statistical approaches (two traditional regression approaches and three causal inference approaches). The results of sensitivity analyses (1) excluding year and (2) excluding meteorological

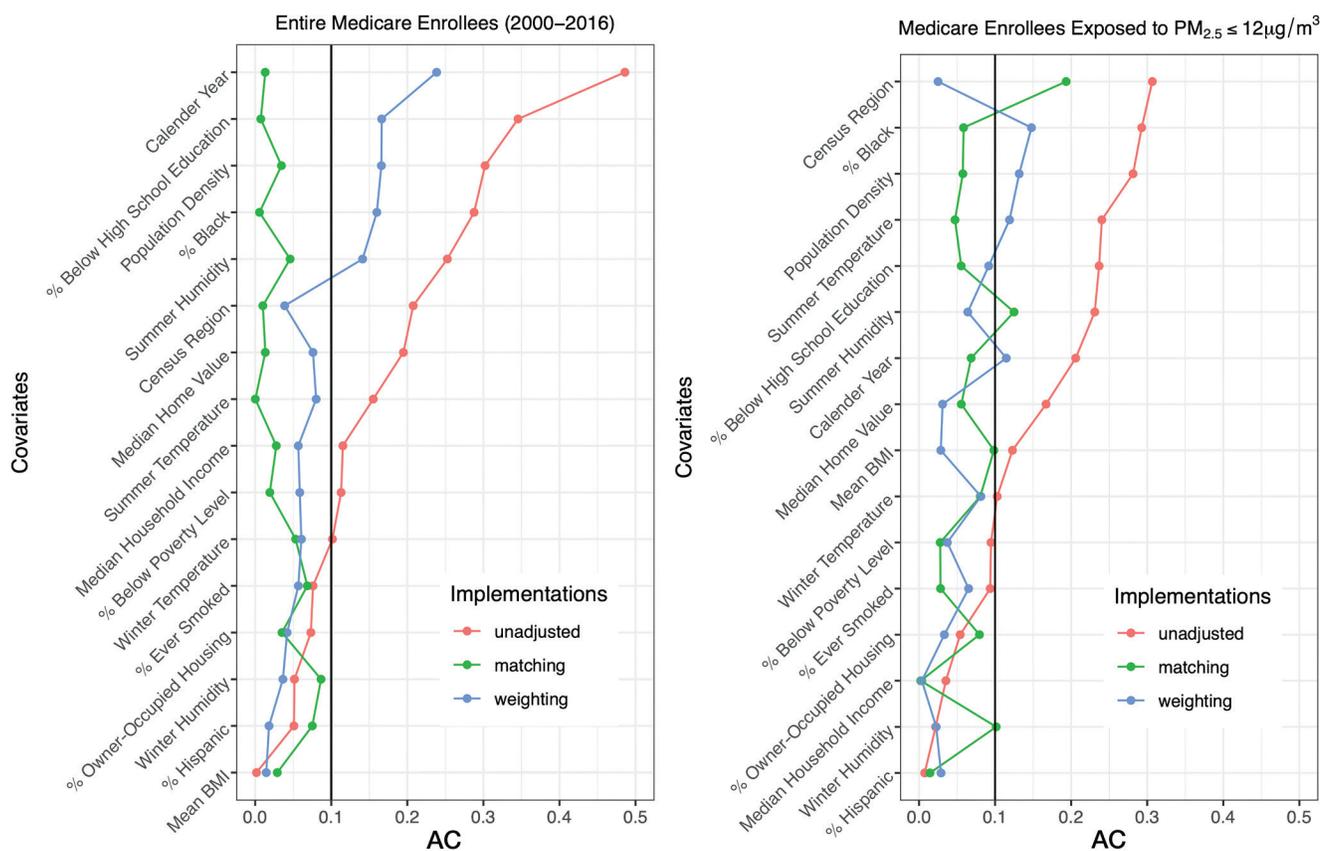


Figure 5. Mean AC for unadjusted, weighted, and matched populations. Mean AC was smaller than 0.1 using causal inference GPS methods (matching and weighting). AC values of <0.1 indicate good covariate balance, strengthening the interpretability and validity of our analyses as providing evidence of causality. From Wu et al 2020*

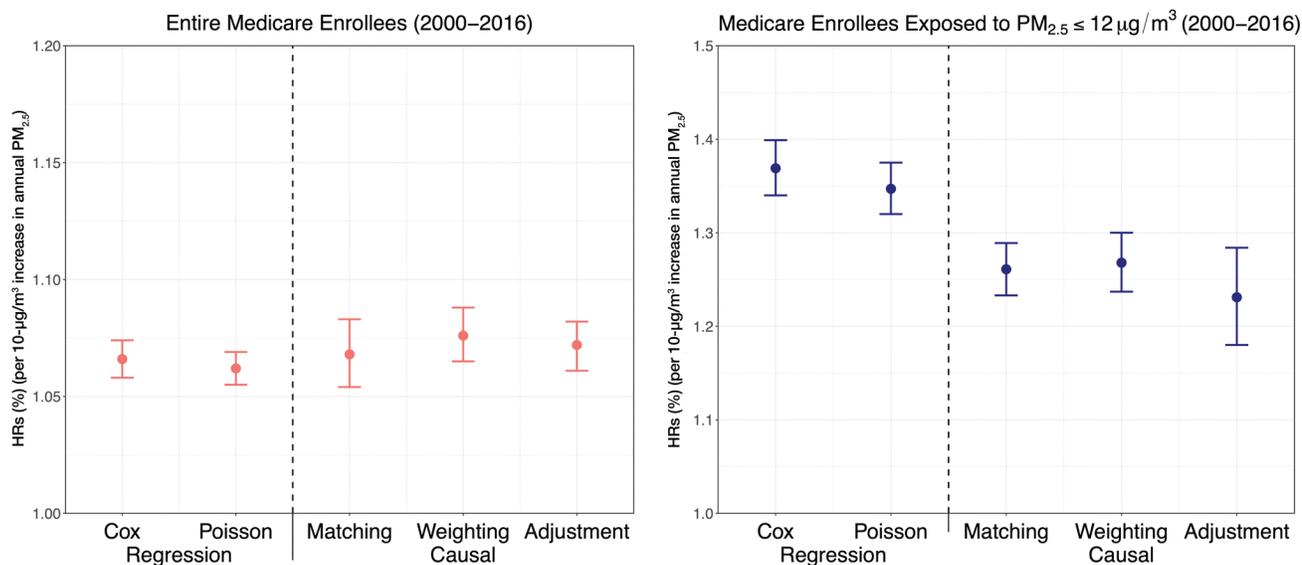


Figure 6. HR and 95% CIs. The estimated HRs were obtained under five different statistical approaches (two traditional approaches and three causal inference approaches) and were adjusted by 10 potential confounders, four meteorological variables, geographic region, and year. From Wu et al 2020*

*Figures 5 and 6 © the Authors, some rights reserved; exclusive licensee AAAS. Distributed under a CC BY-NC 4.0 License, <http://creativecommons.org/licenses/by-nc/4.0/>.

variables are provided in Table 2. When we reanalyzed all data excluding year as a covariate, the estimated HRs were larger in magnitude, potentially indicating residual confounding bias by some unmeasured confounders with time trends that covary with time trends in the outcome and exposure. The estimated HRs for the period 2000–2012 were not identical with those of Di and colleagues (2017b) because of the following updates in our data pipeline: (1) In the Phase 1 report, we used $PM_{2.5}$ exposure data predicted by a hybrid prediction model using a chemical transport model and land-use regression (Di et al. 2016). In this final report, we now use $PM_{2.5}$ exposure data predicted by an ensemble-based model with a better prediction performance (Di et al. 2020). (2) We have reconstructed the principal confounder data from the U.S. Census and American Community Surveys using an updated fully reproducible pipeline described in the Appendix of the report. Given the updates in the data sources, there is no guarantee that the estimated HRs from the statistical models will be identical. Reassuringly, both data analyses found a steady positive relationship between long-term exposure to ambient $PM_{2.5}$ and all-cause mortality and thus further prove that the evidence is robust.

We estimated the total number of deaths avoided among the elderly in a decade if, hypothetically, the U.S. standards had followed the World Health Organization annual guideline of $\leq 10 \mu\text{g}/\text{m}^3$ and all zip codes had complied. For this calculation, we used the most conservative HR estimate across all statistical approaches (HR = 1.06 [95% CI, 1.05 to 1.08] and 1.23 [95% CI, 1.18 to 1.28]). We found that lowering the standards to $10 \mu\text{g}/\text{m}^3$ would have saved 143,257 lives (95% CI, 115,581 to 170,645) in one decade.

Evaluation of Unmeasured Confounding We conducted a sensitivity analysis to evaluate the robustness of our results to unmeasured confounding by calculating the *E*-value. The *E*-value for the point estimate of interest (in our case, the HR) can be defined as the minimal strength of an association, on the risk ratio scale, that an unmeasured confounder would need to have with both the exposure and outcome, conditional on the covariates already included in the model, to fully explain the observed association under the null. We calculated the *E*-values for our reported HRs per $10\text{-}\mu\text{g}/\text{m}^3$ increase of long-term exposure to $PM_{2.5}$. For our main analysis (2000–2016) under a Poisson model, we found that for an unmeasured confounder (*U*) to fully account for (nullify) the estimated effects of the exposure (*E*) on the outcome (*Y*), it would have to be associated with both long-term $PM_{2.5}$ exposure (*E*) and with mortality (*Y*) by a risk ratio of at least 1.32-fold each, through pathways independent of all covariates already included in the model. In other words, if we were to include this *U*, the association between the long-term effects of $PM_{2.5}$ on mortality would become null. A 1.32 risk ratio means that *U* would need to meet the following two criteria: (1) *U* would need to lead to a 32% increase in the risk of mortality (*Y*); and (2) when comparing two groups, one with exposure to

$PM_{2.5}$ that is $10 \mu\text{g}/\text{m}^3$ higher than the other (*E* = low versus *E* = high), the higher exposure group would have a 32% higher prevalence of that unmeasured confounder than the lower exposure group. Please note that the *E*-value cannot address how likely it is for unmeasured confounding to exist. The interpretation of the *E*-value requires substantive knowledge. Also, the *E*-value does not account for potential bias due to other mechanisms, such as measurement error, selection bias, or selective reporting of results. The interpretation of the *E*-value also has to be coupled with other strengths and weaknesses of the study designs (VanderWeele and Ding 2017).

Additional supplementary materials are available online at <https://advances.sciencemag.org/content/suppl/2020/06/26/sciadv.aba5692.DC1>.

Section S1. Statistical Methods

Section S2. Additional Analysis Results

Section S3. Additional Sensitivity Analysis

Section S4. Code

Figure S1. Causal Inference Workflow

Figure S2. ACs, Point Estimates, and 95% CIs of the HRs for the Study Cohort from 2000 to 2012

Figure S3. Standardized Mean Differences (SMDs) for Study Cohort from 2000 to 2016

Figure S4. ACs for Study Cohort from 2000 to 2016 Excluding Year or Meteorological Variables as Confounders in the GPS Model

Figure S5. Estimated Values of GPS

Table S1. Data Sources

Table S2. Characteristics for the Medicare Study Cohort from 2000 to 2012

Table S3. Point Estimates and 95% CIs of the HRs for All Analysis Results

Table S4. The Importance Scores of Variables in the GPS Models

Table S5. *E*-value for Point Estimates and the Lower Bound of the 95% CIs of the HRs for All Analysis Results

4. NATIONAL CAUSAL INFERENCE ANALYSIS ON LONG-TERM EXPOSURE-RESPONSE CURVES FOR $PM_{2.5}$, NO_2 , AND O_3 ON ALL-CAUSE MORTALITY

We applied the proposed GPS matching method to estimate the effect of long-term exposures to $PM_{2.5}$, NO_2 , and O_3 on all-cause mortality. Figure 7 shows the average causal ERCs of each pollutant on all-cause mortality among Medicare beneficiaries (2000–2016). For each pollutant, we present the ERCs in HR associated with long-term exposure to one of the three pollutants with all-cause mortality, using (1) multiple-pollutant models adjusting for the other two pollutants as potential confounders and (2) single-pollutant models without

adjusting for the other pollutants. As a sensitivity analysis, Appendix Figure 1 (available on the HEI Website) shows the ERCs in HRs associated with long-term exposure to $PM_{2.5}$ with all-cause mortality, adjusting for only one pollutant (NO_2) as a potential confounder. We defined the baseline rate as the estimated HR corresponding to an exposure level equal to the 1st percentile of the distribution of that pollutant (the 1st percentile corresponds to 2.77 $\mu\text{g}/\text{m}^3$, 3.41 ppb, and 29.48 ppb for $PM_{2.5}$, NO_2 , and O_3 , respectively). To avoid extrapolation at the support boundaries, we excluded the highest 1% and lowest 1% of pollutants exposures when plotting the ERCs. Therefore, by setting the baseline hazard at the 1% quantile levels, we ensured that $HR = 1.0$ at the starting value of the x-axis for each of the ERCs in all figures.

We found evidence of a harmful causal relationship between mortality and long-term $PM_{2.5}$ exposures adjusted for NO_2 and O_3 across the range of annual averages between 2.77 and 17.16 $\mu\text{g}/\text{m}^3$ (included >98% of observations) in the entire Medicare beneficiaries across the continental United States from 2000 to 2016 (Figure 7). Our results are consistent with recent epidemiological studies reporting a strong association between long-term exposure to $PM_{2.5}$ and adverse health outcomes at low exposure levels. Importantly, the curve is almost linear at exposure levels lower than the current standards, indicating aggravated harmful effects at exposure levels even below the national standards.

There is, in general, a harmful causal impact of long-term NO_2 exposures to mortality adjusted for $PM_{2.5}$ and O_3 across the range of annual averages between 3.4 and 80 ppb (included >98% of observations). Yet at low levels (annual mean ≤ 53 ppb) below the current national standards, the causal impacts of NO_2 exposures on all-cause mortality are nonlinear with statistical uncertainty.

The ERCs of long-term O_3 exposures on all-cause mortality adjusted for $PM_{2.5}$ and NO_2 are almost flat below 45 ppb, which shows no statistically significant effect. Yet we observed an increased hazard when the O_3 exposures were higher than 45 ppb, and the HR is approximately 1.10 when comparing Medicare beneficiaries with annual mean O_3 exposures of 50 ppb with those with 30 ppb.

Comparing the results from multiple- and single-pollutant models, we found that adjusting for the other two pollutants slightly attenuated the causal effects of $PM_{2.5}$ and slightly elevated the causal effects of NO_2 exposure on all-cause mortality. The results for O_3 remained almost unchanged. We also reported results adjusting only for NO_2 (Appendix Figure 1). We found at the low level of $PM_{2.5}$ (below 12 $\mu\text{g}/\text{m}^3$) that the ERC was similar to the ERC obtained when adjusting for both NO_2 and O_3 . However, we found a sharper slope for the ERC in the region with higher levels of $PM_{2.5}$ ($>12 \mu\text{g}/\text{m}^3$).

The multipollutant model results are shown in Appendix Table 1 using both the GPS matching method and multivariate

Poisson regression method, assuming that the ERC is always linear (i.e., a constant HR). We found a consistently statistically significant link between long-term $PM_{2.5}$ exposures and all-cause mortality. Consistent with the ERC results, we found, in general, that adjusting for the other two pollutants slightly attenuated the causal effects for $PM_{2.5}$.

When comparing the ERC results in Figure 7 and the HRs in Appendix Table 1, it is important to note that in the ERC estimate, we used a kernel smoothing approach, a nonparametric approach, to flexibly estimate the ERC. We used the estimated HR at approximately 2.7 $\mu\text{g}/\text{m}^3$ (the 1% quantile of $PM_{2.5}$ levels) as the baseline hazard. In the HR estimate, we used a Poisson regression model (a fully parametric approach). Here we assume a baseline level is at $PM_{2.5}$ levels 0 $\mu\text{g}/\text{m}^3$ and a constant. The regression model coefficient (in Appendix Table 1) represents the HR comparing the mortality rate in the population who were exposed to $PM_{2.5}$ levels that were 10 $\mu\text{g}/\text{m}^3$ higher than those who were exposed to baseline $PM_{2.5}$ levels (0 $\mu\text{g}/\text{m}^3$) under the linearity assumption. Given that the definition of baseline HR (i.e., mortality rate) and the model specifications are different under these two distinct modeling strategies, the HRs were not directly comparable quantitatively. Still, we observed a consistently increased mortality rate linked with the increased $PM_{2.5}$ exposure in the ERC analysis.

We found no evidence of a statistically significant link between long-term NO_2 exposures and all-cause mortality under a GPS matching method with the assumption of a constant HR, whereas we found a statistically significant positive association between long-term NO_2 exposures and all-cause mortality under a multivariate Poisson regression method with the assumption of a constant HR. We also found no evidence of a relationship between long-term O_3 exposures and all-cause mortality under a GPS matching method with the assumption of a constant HR, whereas we found a statistically significant negative association between long-term O_3 exposures and all-cause mortality under a multivariate Poisson regression method with the assumption of a constant HR and adjustment of the other two pollutants.

It is important to note that we found highly nonlinear relationships between both NO_2 and O_3 and all-cause mortality (shown in Figure 7). In these cases, assuming either a GPS matching method or a multivariate Poisson regression method with a constant HR (linear ERC) is subject to model misspecification. Based on the evidence of nonlinearity, we do not believe that these results are correctly modeled when modeled linearly under either the GPS matching method or multivariate Poisson regression method. Such model misspecifications may also partially explain the discrepancy between the results from the GPS matching method and the multivariate Poisson regression method.

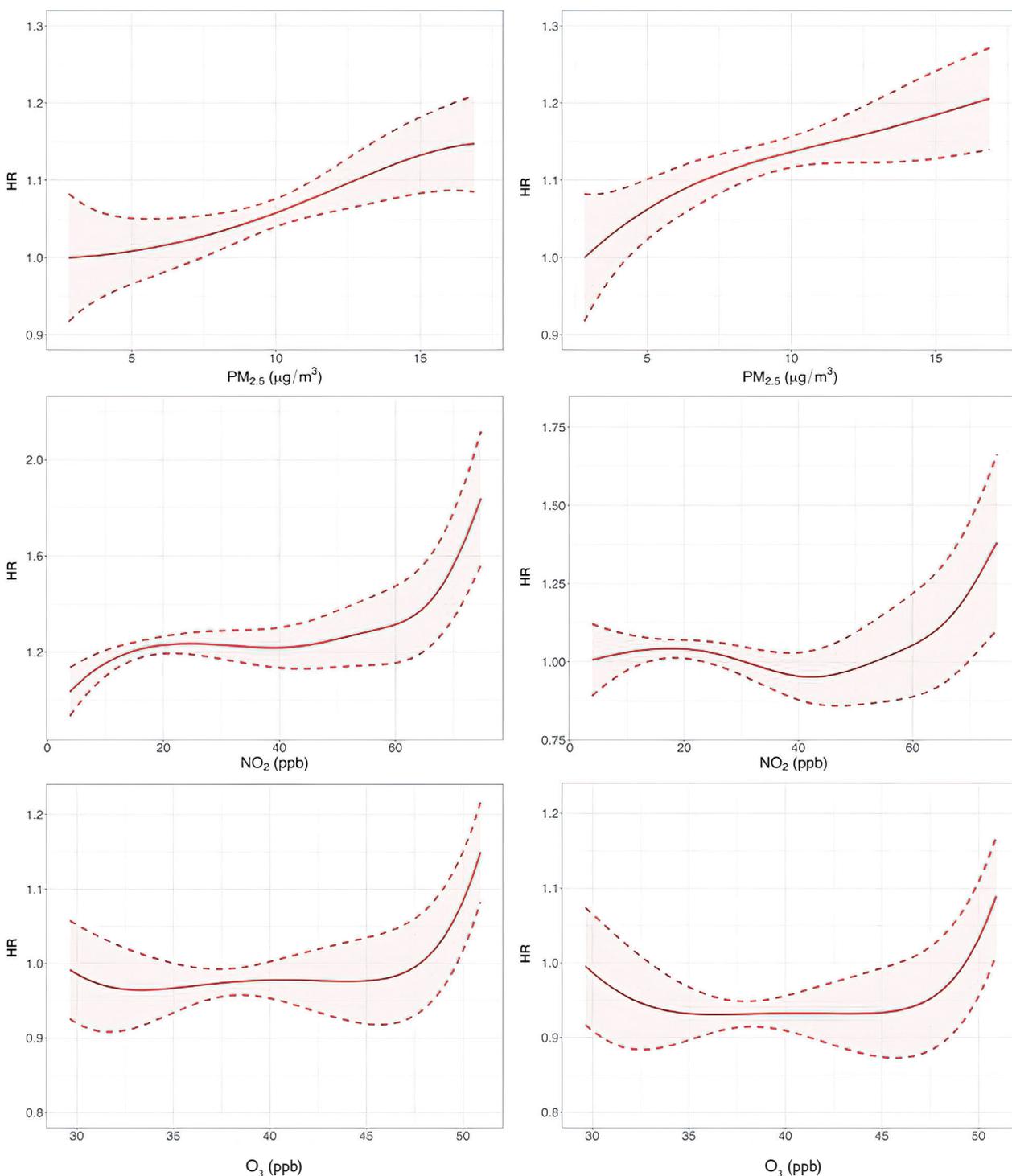


Figure 7. Estimated ERCs relating PM_{2.5}, NO₂, and O₃ to all-cause mortality among Medicare beneficiaries (2000–2016) with associated 95% confidence bands. The left panel presents the ERCs in HRs associating long-term exposure to one pollutant with all-cause mortality adjusting for the other two pollutants as potential confounders. The right panel represents the ERCs of single-pollutant models without adjusting for the other two pollutants. We defined the baseline rate as the estimated hazard rate corresponding to an exposure level set at the 1st percentile of the distribution of each pollutant. The HRs were calculated as the ratio of the hazard rate at every exposed level to the baseline rate. To avoid potential unstable behavior at the support boundaries, we excluded the highest 1% and lowest 1% of pollutants exposures.

5. PIPELINE FOR REPRODUCIBLE RESEARCH

In this section, we present our open science framework for the largest study conducted to date on the long-term effects of air pollution on mortality. Before we present details on our pipeline for reproducibility, we start with a general introduction to challenges and opportunities in reproducible research.

5.1. INTRODUCTION

Threats of high costs associated with implementation and compliance with air quality regulations have spurred increasingly contentious legal challenges to these regulations, and the scientific evidence for harmful effects of air pollution is being subjected to unprecedented scrutiny. Central to current debates are issues related to access to the data, transparency, and reproducibility of studies that constitute the scientific basis supporting regulatory decisions.

The Clean Air Act requires the EPA to periodically review the science for six major air pollutants, including particulate matter. The EPA's National Center for Environmental Assessment develops Integrated Science Assessments summarizing the science related to the health and ecological effects caused by these six pollutants. Integrated Science Assessments provide a comprehensive review of the policy-relevant scientific literature published since the last NAAQS review and are a critical part of the scientific basis for establishing the NAAQS.

5.2. REPLICABILITY VERSUS REPRODUCIBILITY

A study is *replicated* when *new data* are collected and analyzed, independently, by a new set of investigators (Stodden et al. 2016). Air pollution studies in the United States, Europe, and countries around the world have provided consistent evidence that exposure to $PM_{2.5}$ increases the risk of death and other adverse health outcomes. Consistency of the evidence across many studies is a step toward independent *replication*.

A study is *reproduced* when the *same data* are reanalyzed, independently, by a new set of investigators (Peng 2015). Classic examples of reproducibility were the reanalyses of the Harvard Six Cities Study and the American Cancer Society Study (Krewski et al. 2003; Krewski et al. 2009). However, most air pollution health studies are not *reproducible*, often because of the strict privacy requirements these studies must abide by. The historical reliance on cohort health studies creates barriers to reproducibility that are nearly impossible to overcome. Because of the privacy restrictions on health data collected on the participants in such cohort studies, the underlying data simply cannot be shared: redacting “names, addresses, any other identifying information” from confidential data, as has been previously advocated, is not enough to make it sharable. Furthermore, reproducing and updating results from such studies with more recent data is challenging because these cohorts are typically “closed” for

further enrollment of participants. Initiating new cohort studies is prohibitive because of the sheer cost of conducting such studies and the length of time—often 10 to 20 years—before results can be obtained.

Despite the reanalyses of the Harvard Six Cities Study and the American Cancer Society Study, as well as the open process used by the EPA to review all the scientific evidence (U.S. Environmental Protection Agency 2020b), there is still the challenge of going beyond consistent evidence. We argue that in this new era of data science, although in principle it is possible to ensure full reproducibility, the process of doing so is highly complex because of the necessary sophistication of the data and analytical methods.

5.3. TOWARD OPEN SCIENCE IN AIR POLLUTION EPIDEMIOLOGY: A FIRST

In a recent article of ours (Wu et al. 2020), we reported on the largest study conducted to date on the long-term effects of air pollution ($PM_{2.5}$) on mortality (also see Section 2 of this report). Here we provide the details that make this huge study *reproducible*. We stress that the study relied entirely on publicly available data. More specifically, to overcome the privacy barriers to transparency and reproducibility, we did not use a traditional cohort. We relied instead on *privacy-protected but publicly available* Medicare health data that included almost 97% of the U.S. population older than 65 years over the years 2000–2016. We have no influence on who has access to the Medicare health claims data we used in our study; any individual or institution can submit a data access request to the Centers for Medicare & Medicaid Services and obtain the exact same files we used (see <https://www.resdac.org/> for details on how to obtain these files).

In most cohort studies, the data processing and statistical software programs to implement analyses are not generally made publicly available. In Wu and colleagues (2020), we made the software code and workflows available in open, trusted digital repositories. Reproducibility instructions and open-source software are hosted on GitHub and are publicly available at <https://github.com/NSAPH/National-Causal-Analysis>. In Figure 8, we describe the steps of the workflow we used in Wu and colleagues (2020).

We begin the “Data Acquisition” step (Figure 8, left panel) by curating and acquiring data sets from a diverse set of sources. These included publicly purchasable health data obtained from the Centers for Medicare & Medicaid Services (<https://www.resdac.org/>; we purchased the following file: 100% MBSF - MASTER BENEFICIARY SUMMARY FILE [MBSF] base, years); freely available data on population-level characteristics obtained from the Centers for Disease Control and Prevention and the United States Census Bureau; and freely available meteorological data from Google Earth Engine. In addition, freely available data from the National Aeronautics and Space Administration, National Oceanic and

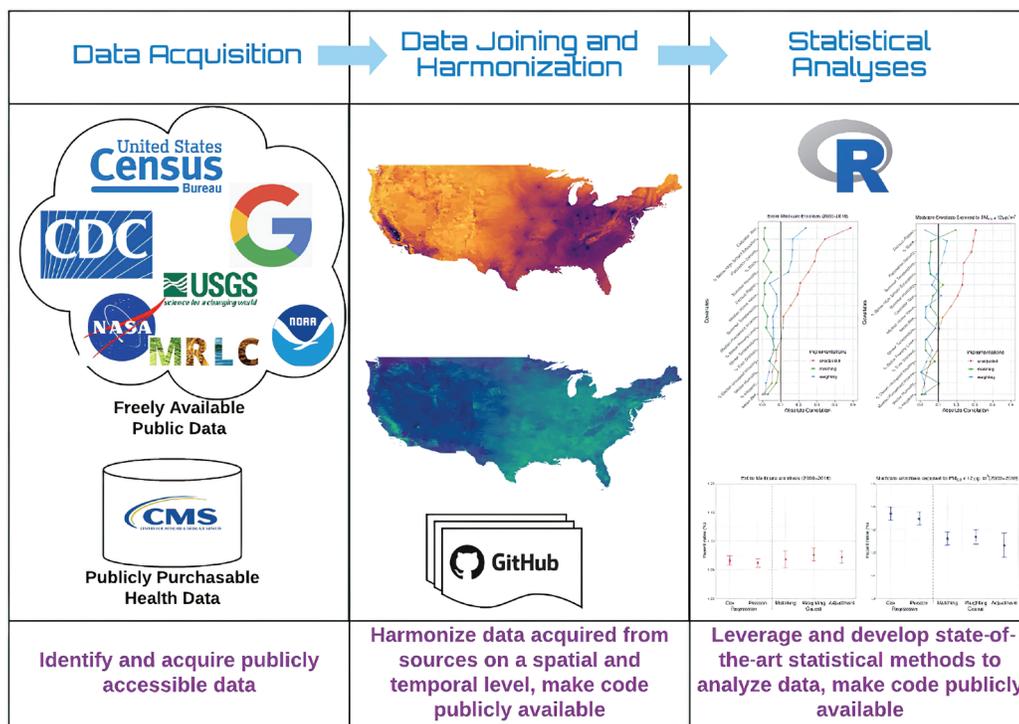


Figure 8. High-level overview of our data pipeline. The data pipeline involves three steps: (1) “Data Acquisition” (left panel), which involves identifying and acquiring publicly accessible data for the analysis; (2) “Data Joining and Harmonization” (middle panel), which involves developing and applying code for the processing of the data to ensure harmonization across data sources with regard to spatial and temporal resolutions; and (3) “Statistical Analyses” (right panel), which involves the development and application of novel statistical methodology to analyze the data. Code for all steps has been made publicly available on GitHub to ensure reproducibility (<https://github.com/NSAPH/National-Causal-Analysis>). (Statistical Analyses figures from Wu et al. 2020, © the Authors, some rights reserved; exclusive licensee AAAS. Distributed under a CC BY-NC 4.0 License, <http://creativecommons.org/licenses/by-nc/4.0/>.)

Atmospheric Administration, United States Geological Survey, and Multi-Resolution Land Characteristics Consortium were used to develop our air pollution prediction models. Additional details on the data acquisition can be found in the GitHub repository.

Following acquisition of the data, the “Data Joining and Harmonization” step (Figure 8, middle panel) involved developing and applying code for data processing to ensure that these diverse sources could be used together, by linking and harmonizing the data based on spatial and temporal resolutions. All code used for the data joining and harmonization is made publicly available on GitHub to ensure reproducibility.

The last step involves the “Statistical Analyses” (Figure 8, right panel), for which we developed novel statistical methodology to analyze the data. All code used for the statistical analyses is made publicly available on GitHub as well (<https://github.com/NSAPH/National-Causal-Analysis>). Table 3 lists all of the publicly available original data sources used in our analyses.

5.4. DETAILS ON THE PIPELINE FOR ENSURING REPRODUCIBILITY

To ensure the reproducibility of our workflow, we developed software codes that allow investigators to reproduce the entire pipeline associated with our study “Evaluating the Impact of Long-Term Exposure to Fine Particulate Matter on Mortality Among the Elderly” (Wu et al. 2020). This code is openly available on GitHub (<https://github.com/NSAPH/National-Causal-Analysis>). Reproducible code is provided to run all analysis after obtaining the publicly available exposure data (<https://beta.sedac.ciesin.columbia.edu/data/set/aqdh-pm2-5-concentrations-contiguous-us-1-km-2000-2016>) and after purchasing the following Centers for Medicare & Medicaid Services file for each year for the health data: 100% MBSF.

More specifically, the GitHub repository is structured into five components, described in detail below:

1. **Confounders Directory** — Contains the process and code by which the zip code-level demographic data, smoking and BMI data, and weather data are acquired and prepared for use, including details on how to acquire them

Table 3. Data Sources^a

Source	Dataset	Website
NOAA	Reanalysis meteorological data	http://www.noaa.gov/
NASA	MAIAC AOD data	
	Surface reflectance data	https://www.nasa.gov/
	NDVI data	
	OMI Aerosol Index Data GEOS-Chem simulation outputs	http://acmg.seas.harvard.edu/geos/
U.S. Geological Survey	Global terrain elevation data	https://lta.cr.usgs.gov/
Census Bureau	Road density, population count and area	https://www.census.gov/
MRLC	National Land Cover Dataset	https://www.mrlc.gov/
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/

AOD = aerosol optical depth; GEOS = Goddard Earth Observing System; MAIAC = Multi-Angle Implementation of Atmospheric Correction; MRLCD = Multi-Resolution Land Characteristics (consortium); NASA = National Aeronautics and Space Administration; NDVI = normalized difference vegetation index; NOAA = National Oceanographic and Atmospheric Administration; OMI = Ozone Monitoring Instrument.

^a Detailed list and software codes are available at <https://github.com/NSAPH/National-Casual-Analysis>.

2. **Exposures Directory** — Describes the preparation of the PM_{2.5} data and links to them
3. **HealthOutcomes Directory** — Contains the code used to process data from the Medicare Beneficiary Summary Files obtained from the Centers for Medicare & Medicaid Services
4. **MergedData Directory** — Contains the process and code by which all these data sources can be combined for analysis
5. **StatisticalAnalysis Directory** — Contains code for conducting all statistical analyses reported in the current report

We have included as much data as we are allowed to share and can feasibly include in a GitHub repository (some files are too large to share). Where we were unable to share data, we have provided instructions on how to acquire the source data and prepare it for use with the data pipelines.

Confounders Directory The confounders directory contains the code for preparing data used as confounders, with the following subdirectories. For each subdirectory, we have described the input data, provided processing code, and described the final output.

1. The BRFSS directory contains the code needed to process smoking and BMI information from the Centers for Disease Control and Prevention's BRFSS data set. All data have been linked from the county level to the zip code level, as shown in Appendix Figure 2 (available on the HEI Website).
2. The census directory contains the code needed to process zip code-level demographic data from the U.S. Census and American Community Survey at the zip code level, as shown in Appendix Figure 3 (available on the HEI Website).
3. The earth_engine directory contains the code needed to process zip code-level temperature, humidity, and precipitation data from Google Earth Engine, as shown in Appendix Figure 4 (available on the HEI Website).

Exposures Directory For the Exposures Directory, we used a series of .RDS files containing annual estimates of mean PM_{2.5} for each zip code (Di et al. 2019). These are available publicly at <https://beta.sedac.ciesin.columbia.edu/data/set/aqdh-pm2-5-concentrations-contiguous-us-1-km-2000-2016>. We have included a description of how the data were created and a link to the data.

HealthOutcomes Directory The HealthOutcomes directory contains the code needed to prepare the initial mortality data set before merging it with the confounder and exposure data. We used the Medicare Beneficiary summary file from 1999–2016 to create this data set.

MergedData Directory The MergedData directory contains the code needed to clean and merge exposure, covariate, and health data to produce combined data sets covering the period 1999–2016 that can be used to estimate the effects of air quality exposures on health outcomes. This highly detailed process is shown in Appendix Figure 5 (available on the HEI Website), the output of which is the dataset used for our statistical analysis.

StatisticalAnalysis Directory The StatisticalAnalysis directory contains the code needed to conduct the statistical analysis. The input is the merged fst files described in the MergedData directory.

5.4.1. Computational Aspects

All processing and analysis described above were conducted on the Research Computing Environment (Hammer et al. 2020) supported by the Institute for Quantitative Social Science in the Faculty of Arts and Sciences at Harvard University. It should be noted that the scale of the Research Computing Environment in terms of computing and storage resources allowed us to undertake studies on the entire Medicare population, as described above. This study is *reproducible*: it relies entirely on publicly available data, as described above. In a recent commentary in *Science*, Susan Cosier (2018) pointed to the importance of our work for promoting open, reproducible evidence that can be used to inform public policy.

6. ONGOING WORK

In this section, we summarize our ongoing work on the project.

6.1. HARMONIZED ANALYSES ACROSS U.S., CANADIAN, AND EUROPEAN COHORTS

We are grateful to HEI for allowing us to continue our analyses for one more year with the specific goal of increasing harmonization across three key studies on low-level exposure: (1) the Harvard Medicare study in the United States (PI Francesca Dominici); (2) the Mortality–Air Pollution Associations in Low-Exposure Environments (MAPLE) study in Canada (PI Michael Brauer); and (3) and the Effects of Low-Level Air Pollution: A Study in Europe (ELAPSE) study (PI Bert Brunekreef). As part of our proposed work, we will do the following:

1. Identify common analyses, using, for example, similar statistical methods, similar spatial resolution for the exposure models, and a common set of confounders across the studies.

2. Develop a professional R software package and facilitate the application of our methods for causal inference to Canadian and European cohorts.
3. Assess the sensitivity of our exposure effect estimates to various adjustments for confounding and to various approaches to spatial aggregation. Our models will include the same sets of confounders being investigated by the other two research groups, depending on data availability. Medicare data has a limited number of individual-level confounders but an extensive number at the zip code level.
4. Apply the extended shape constrained health impact function approach (Brauer et al. 2019) to the Medicare data with a harmonized set of covariates to estimate a causal response function. This will provide the opportunity to conduct a harmonized analysis that will use a common set of covariates and a common approach for the estimation of the ER function across all three cohorts.

Repeat our previous analysis using the following exposure estimates from the MAPLE study (Randall Martin’s team [<https://sites.wustl.edu/acag/datasets/>]):

- a. Surface $PM_{2.5}$ + Components – 2000–2017 (van Donkelaar et al. 2019 #83)
- b. Surface NO_2 – 1996–2012 (Geddes et al. 2016 #111)

6.2. SOFTWARE DEVELOPMENT

We have established a collaboration with professional software engineers Mahmood Shad and Naeem Khoshnevis (Faculty of Arts & Sciences Research Computing, Harvard University, Cambridge, MA) to refactor the source code of a paper by Wu and colleagues (in review) and create an R package, meeting the best software engineering practices. This will allow us to

1. Provide an important tool for the research community.
2. Overcome computational scalability issues.
3. Allow the ELAPSE team to use the code on their own datasets with few or no barriers. As part of the code development, we will ask the ELAPSE team to share synthetic data with us to ensure that the code is refactored to meet a wide range of input datasets, including their data.

7. DISCUSSION AND CONCLUSIONS

Our work provides comprehensive evidence about causal associations between exposure to $PM_{2.5}$, NO_2 , and O_3 and various health outcomes. More specifically, in this final report, we have reported results on the causal link between long-term exposure to $PM_{2.5}$, even at $PM_{2.5}$ levels below or equal to $12 \mu g/m^3$ and mortality among Medicare beneficiaries (Wu et al. 2020). This work relies on newly developed causal inference methods for continuous exposures (Wu et al. in review).

Considering (1) the massive study population, (2) the numerous sensitivity analyses, and (3) the transparent assessment of covariate balance that indicates the quality of causal inference for simulating randomized experiments, we conclude that conditionally on the required assumptions for causal inference, collectively our results indicate that long-term $PM_{2.5}$ exposure is likely to be causally related to mortality. This conclusion assumes that the causal inference assumptions hold and, more specifically, that we were able to adequately account for confounding bias. We explored various modeling approaches and conducted extensive sensitivity analyses and found the results were robust across approaches and models. The work relied on publicly available data, and we have provided code that allows for reproducibility of our analyses. The understanding of underlying causal mechanisms among various pollutants in atmospheric chemistry is still evolving. Whether ambient NO_2 or O_3 serve as confounding factors in the relationship between ambient $PM_{2.5}$ and health outcomes (including mortality) is still an open question. To maintain the maximum scientific rigor, we drew our conclusions from single-pollutant models. We conducted multipollutant analyses to assess the degree to which other pollutants could potentially alter our conclusions. Reassuringly, our final conclusions about the link between long-term exposure to $PM_{2.5}$ and increased all-cause mortality in the Medicare population at low levels of ambient $PM_{2.5}$ remain unchanged, although the effect size was attenuated.

Both traditional and causal inference approaches rely on assumptions. Unless all assumptions are satisfied, regardless of approach, demonstration of causal effects is not guaranteed. A critical assumption that guarantees our conclusion's validity is that our statistical analyses accounted for all confounders. This assumption must always be made in observational studies. We accounted for individual- and area-level potential confounders. To mitigate unmeasured confounding bias, we assessed the results' sensitivity by including year as a surrogate for some unmeasured confounders that might have covaried over time with $PM_{2.5}$ and mortality and thus confounded their association. Even after adjustment for year, the analysis could still have been affected by confounding bias by unmeasured factors. We therefore conducted further sensitivity analyses to unmeasured confounding by calculating the E -value and showed that our results were robust to unmeasured confounding bias.

Dominici and Zigler (2017) previously discussed three notions of what constitutes evidence of causality in air pollution epidemiology. The first is causality inferred from evidence of biological plausibility (Stanek et al. 2011). The second is consistency of results across many epidemiological studies and adherence to Bradford Hill causal criteria (Hill 1965). The third is the use of causal inference methods that are more robust to model misspecification compared with traditional approaches and that, when assumptions are

met, can isolate causal relationships. More specifically, the causal inference approaches considered in the current work require the estimation of a GPS as the first step. Assuming all causal inference assumptions hold, these approaches are more robust to outcome model misspecification and allow for the transparent evaluation of covariate balance. However, it is important to note that if the models are accurately specified and all assumptions are met, traditional approaches have the potential to help identify causal relationships as well.

In the case of analysis at low levels of $PM_{2.5}$ (below or equal to $12 \mu\text{g}/\text{m}^3$), it is more likely that the traditional methods are subject to model misspecification and thus that the results may be biased. Both methods provide meaningful scientific evidence that higher $PM_{2.5}$ levels are linked to higher risks of all-cause mortality, given that the underlying statistical assumptions were met. We have conducted additional sensitivity analyses using a Poisson model in which we added penalized spline terms for every potential confounder, thus allowing for flexible nonlinear adjustments. We found that a more flexible regression model specification may help adequately adjust for confounding; when implementing these models, we found results similar to those for the causal inference approaches (Appendix Table 2). However, running multivariate regression models with flexible splines for every potential confounder is much more computationally burdensome. In addition, newly developed causal inference methods allow a transparent assessment of covariate balance. The covariate balance assessment can help researchers understand whether their models are adequately controlled for every measured confounder. Such assessment is not straightforward in traditional multivariate regression approaches.

Our work estimates causal relationships using causal inference methods, addressing just one of Dominici and Zigler's three notions of what constitutes scientific evidence of causality. The collective evidence across studies conducted in various populations, using various study designs and methods, is also imperative to inform regulatory action. A recent meta-analysis found robust evidence for an effect on mortality across 52 cohort studies at $PM_{2.5}$ levels below $10 \mu\text{g}/\text{m}^3$ (Vodanos et al. 2018).

Exposure to $PM_{2.5}$ was estimated from a prediction model, which, while very good, was not perfect. The $PM_{2.5}$ exposure prediction model developed by Di and colleagues (2019) that was used in this analysis indicated excellent model performance, with a 10-fold cross-validated R^2 of 0.89 for annual $PM_{2.5}$ predictions. However, exposure error could have affected all the HR estimates. In the original study by Di and colleagues (2017b), the investigators assessed the robustness of the results to the exposure predictions by repeating the analysis based on $PM_{2.5}$ exposure data obtained from 1,928 EPA ambient monitors. The additional analysis was restricted to the subpopulation of individuals within 50 km of these monitors. Although this subset did not represent the entire

population, we found that the analysis based on nearest monitoring site led to an HR estimate that was only slightly lower than the one obtained using the exposure prediction model (i.e., 1.061 [95% CI, 1.059 to 1.063] versus 1.073 [95% CI, 1.071 to 1.075]). Although these results are reassuring, we recognize that they are not a substitute for a formal analysis that accounts for exposure error. Furthermore, in Section 2, we provided evidence that the accuracy of these exposure prediction models for $PM_{2.5}$ was actually higher at lower concentrations.

How to propagate exposure error under a causal inference framework for a continuous exposure under a causal inference framework is still an area of active research; the presence of exposure measurement error could induce a bias toward the null in all of our estimates (Kioumourtoglou et al. 2014). The majority of causal inference methods make the simplifying assumption of an exposure measured without error (Bang and Robins 2005; Hernán et al. 2000; Robins et al. 2000; Rosenbaum and Rubin 1983; Rubin and Thomas 1996; Van der Laan and Rose 2011). The issue of an error-prone exposure drastically complicates causal inference problems. Failure to account for exposure error has been shown to lead to invalid inference (Carroll et al. 2006; Sarnat et al. 2010; Szpiro et al. 2011). To our knowledge, methods for estimating causal ERCs that account for error-prone exposures do not exist. Although measurement error has been extensively studied outside of causal inference settings, accounting for exposure measurement error in causal inference for continuous exposures is a completely new endeavor. This is because the exposure measurement error affects (1) estimation of the GPS, (2) implementation of the GPS (e.g., matching and weighting), and (3) the health outcome model (Braun et al. 2017). In addition, in the context of our studies, where air pollution exposures were aggregated to zip codes (because exact residential addresses were not available), additional uncertainty may arise from the aggregation procedure. The very limited literature addressing error-prone exposures in causal inference is confined to binary and categorical exposures (Babanezhad et al. 2010; Braun et al. 2017; Wu et al. 2019).

This is an active area of research, and we are developing approaches that will propagate exposure error into causal estimates of health effects for the entire Medicare population. Regression calibration is a common method for measurement error correction (Carroll et al. 2006). Wu and colleagues (2019) proposed a regression calibration approach for GPS analysis under categorical exposures. The proposed approach was applied in the context of long-term $PM_{2.5}$ exposure and mortality using Medicare data in the Northeastern United States. When accounting for exposure error, there was a stronger and statistically significant association between exposure to $PM_{2.5}$ and mortality, although with larger CIs. We are working to overcome the computational bottlenecks that would allow us to extend this approach to the realm of continuous

exposures. At the same time, we are also exploring alternative approaches, including the extension of the previously developed framework to a Bayesian framework.

In addition, it is also important to account for potential measurement error in covariates. Methodological work on how to propagate error in covariates under a causal inference framework is at its infancy (see, for example, Hong et al. 2019; McCaffrey et al. 2013; Steiner et al. 2011; Stürmer et al. 2005; Webb-Vargas et al. 2017). In our context, obtaining validation data to try to adjust for measurement error in confounders is challenging. Additionally, integrating measurement error in both exposure and confounders would require the development of new statistical methods and is the subject of future work.

Our model parameterization assumes that zip code-specific information is spatially independent, given covariates. Because we adjusted for numerous zip code-level predictors of mortality, including SES and meteorological variables, this assumption is likely to hold. If any residual spatial dependence remains under certain assumptions (e.g., those used in generalized estimating equations), it would not have affected our point estimates but could have influenced the estimated standard errors. However, our bootstrapping procedure partially accounts for this possibility. By randomly sampling zip codes for each bootstrap replicate, we were able to break down spatial dependence, given covariates. Therefore, it is unlikely that our results were affected by spatial correlation.

We adjusted for potential spatial confounding that was not captured by zip code-level observed covariates by including a dummy region variable. Covariate balance was achieved across all variables, including the region variable, using the entire Medicare population; the absolute correlations are <0.1 for every observed covariate. The absolute correlation was larger than 0.1 for the region covariate in the low-level exposure subset. However, the absolute correlation for the region covariate was still relatively small (<0.2) and has been largely improved compared with the absolute correlation calculated in the unadjusted observed data. However, given the imbalance of region in the analysis on the cohort of Medicare beneficiaries that were always exposed to $PM_{2.5}$ levels lower than $12 \mu\text{g}/\text{m}^3$ (i.e., the low-level exposure subset), potential spatial confounding may still be a concern. In future work, we plan to use exact matching to match on the geographic region variable to further improve the balance of the matched set on the spatial confounder. In the exact matching, we only allowed the comparison between matched pairs that belonged to the same geographic region. We will conduct region-specific analyses to evaluate whether the relationship between exposure to $PM_{2.5}$ and all-cause mortality varies throughout regions. We will also incorporate dummy spatial variables with finer geographic resolutions into the models.

We acknowledge that a big limitation of all our analyses is the hybrid nature of the study design. Medicare claims are available at the individual level, and they include information

on age, sex, race, and eligibility for Medicaid (a proxy for low income). As health outcomes, Medicare claims include individual-level information on cause-specific hospitalizations and all-cause mortality. The ideal design would allow us to include information on the geocoded address of each Medicare beneficiary as well as information on a very extensive list of individual-level potential confounders, such as smoking, BMI, and socioeconomic variables. Unfortunately, this information is not available. Because residential addresses are only available at the zip code level, we were obliged to aggregate the air pollution exposure levels from 1-km \times 1-km grids to the zip code and assign the same exposure to all the Medicare beneficiaries living in the same zip code. In our current and future work, we are planning to address at least two sources of exposure error: (1) one deriving from the fact that exposures were estimated and (2) another deriving from the fact that the exposures must be aggregated at the zip code level.

An additional problem of the hybrid nature of our study design was that, except for race, sex, and dual eligibility to Medicare and Medicaid, information on all the other potential confounders was available at zip code level only. Furthermore, we could only adjust for smoking at the county level, and we recognize that this is less than ideal. To increase confidence in our results, we conducted a study by Makar and colleagues (2017), in which we linked Medicare claims data to data from the MCBS at the individual level. The MCBS provides information on an extensive list of individual-level behavioral risk factors (over 100 potential measured confounders at the individual level). This extensive list includes patients' functional status (e.g., if they have difficulty walking), their behavioral risk factors (e.g., smoking status), and their detailed demographics (e.g., marital status and level of education), among others. For all of the outcomes we examined (all-cause mortality, cardiovascular hospitalization, and respiratory hospitalizations), we found that the estimated HRs remained unchanged when we excluded the MCBS variables among the confounding variables used for the adjustment.

However, although we acknowledge the potential limitations in our hybrid design, it is important to note that for the Poisson and causal inference models our unit of analysis was counts of individuals at the zip code level in a given year, not individuals. Our unit of analysis, air pollution exposures, confounders, and counterfactuals were all at the zip code level. The hybrid design allowed us to stratify by some individual-level variables (e.g., sex, race, age, and Medicaid eligibility) while accounting for zip code-level confounders in the models.

Our studies have been based on publicly available data sources, and we have made all code developed for our analyses publicly available. Our approach maximizes reproducibility and transparency. We provide robust evidence that the current U.S. standards for $PM_{2.5}$ concentrations are not protective enough and should be lowered to ensure that vulnerable populations, such as the elderly, are safer.

Our results raise awareness of the continued importance of assessing the impact of air pollution exposure on mortality. There are currently numerous disputes about the evidence from previous air pollution epidemiological studies, with arguments made for only using causal inference methods or only including studies that make participants' information publicly available. We strongly oppose these. Most epidemiological studies rely on confidential patient data to provide evidence on adverse health effects of environmental exposures and focus on populations that cannot be studied using administrative data. We hope this work will help researchers and policy makers, particularly as discussions of revising national $PM_{2.5}$ standards are underway.

ACKNOWLEDGMENTS

We would like to thank HEI for its support of our work, without which we would not have been able to advance our methods or evidence base at the scope and scale that are 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 reports 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, Ph.D., ELS (science writer and editor, The Tobin Touch, Arlington Heights, IL), and Leila Kamareddine, MPH (Program Coordinator, Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA), who provided support in the preparation of this report.

REFERENCES

- Abatzoglou JT. 2013. Development of gridded surface meteorological data for ecological applications and modeling. *Int J Climatol* 33:121–131.
- Andersen PK, Gill RD. 1982. Cox's regression model for counting processes: A large sample study. *Ann Statist* 10:1100–1120.
- Austin PC. 2019. Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Stat Methods Med Res* 28:1365–1377.
- Babanezhad M, Vansteelandt S, Goetghebeur E. 2010. Comparison of causal effect estimators under exposure misclassification. *J Stat Plan Inference* 140:1306–1319.

- Bang H, Robins JM. 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics* 61:962–973.
- Brauer M, Brook JR, Christidis T, Chu Y, Crouse D, Erickson A, et al. 2019. Mortality–air pollution associations in low-exposure environments (MAPLE): Phase 1. Available: www.healtheffects.org/publication/mortality%E2%80%93air-pollution-associations-low-exposure-environments-maple-phase-1 [accessed 13 January 2021].
- Braun D, Gorfine M, Parmigiani G, Arvold ND, Dominici F, Zigler C. 2017. Propensity scores with misclassified treatment assignment: A likelihood-based adjustment. *Biostatistics* 18:695–710.
- 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.
- Carone M, Dominici F, Sheppard L. 2020. In pursuit of evidence in air pollution epidemiology: The role of causally driven data science. *Epidemiology* 31:1–6.
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective*, second edition. Boca Raton, FL: CRC Press.
- Centers for Medicare & Medicaid Services. Available: <http://www.cms.gov> [accessed 1 November 2017].
- Chen T, Guestrin C. 2016. Xgboost: A scalable tree boosting system. In *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13–17 August 2016, San Francisco, CA. Available: <https://doi.org/10.1145/2939672.2939785> [accessed 27 June 2020].
- Cosier S. 2018. Clever use of public data could sidestep new rule. *Science* 360:473.
- Crouse DL, Peters PA, Hystad P, Brook JR, van Donkelaar A, Martin RV, et al. 2015. Ambient PM_{2.5}, O₃, and NO₂ 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.
- 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.
- Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, et al. 2019. An ensemble-based model of PM_{2.5} concentration across the contiguous United States with high spatiotemporal resolution. *Environ Int* 130:104909.
- Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, et al. 2020. Assessing NO₂ concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. *Environ Sci Technol* 54:1372–1384.
- 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.
- Di Q, Koutrakis P, Schwartz J. 2016. A hybrid prediction model for PM_{2.5} mass and components using a chemical transport model and land use regression. *Atmospheric Environ* 131:390–399.
- Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. 2017b. Air pollution and mortality in the Medicare population. *N Engl J Med* 376:2513–2522.
- 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.
- Dominici F, Bargagli-Stoffi FJ, Mealli F. 2020. From controlled to undisciplined data: Estimating causal effects in the era of data science using a potential outcome framework. Available: <https://arxiv.org/abs/2012.06865> [accessed 13 January 2021].
- 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. Available: www.healtheffects.org/publication/assessing-adverse-health-effects-long-term-exposure-low-levels-ambient-air-pollution [accessed 27 June 2020].
- Dominici F, Zigler C. 2017. Best practices for gauging evidence of causality in air pollution epidemiology. *Am J Epidemiol* 186:1303–1309.
- Geddes JA, Martin RV, Boys BL, van Donkelaar A. 2016. Long-term trends worldwide in ambient NO₂ concentrations inferred from satellite observations. *Environ Health Perspect* 124:281–289.
- Goldman GT, Dominici F. 2019. Don't abandon evidence and process on air pollution policy. *Science* 363:1398–1400.
- Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens Environ* 202:18–27.
- Hales S, Blakely T, Woodward A. 2012. Air pollution and mortality in New Zealand: Cohort study. *J Epidemiol Community Health* 66:468–473.
- Hammer MS, van Donkelaar A, Li C, Lyapustin A, Sayer AM, Hsu NC, et al. 2020. Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environ Sci Technol* 54:7879–7890.
- Haneuse S, VanderWeele TJ, Arterburn D. 2019. Using the E-value to assess the potential effect of unmeasured confounding in observational studies. *JAMA* 321:602–603.

- Hernán MA, Brumback B, Robins JM. 2000. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 11:561–570.
- Hill AB. 1965. The environment and disease: Association or causation? *Proc R Soc Med* 58:295–300.
- Hirano K, Imbens GW. 2004. The propensity score with continuous treatments. In: *Applied Bayesian Modeling and Causal Inference from Incomplete-data Perspectives*, (Gelman A, Meng X-L, eds). Hoboken, NJ: John Wiley & Sons, Ltd.
- Hong H, Aaby DA, Siddique J, Stuart EA. 2019. Propensity score-based estimators with multiple error-prone covariates. *Am J Epidemiol* 188:222–230.
- Imai K, Ratkovic M. 2014. Covariate balancing propensity score. *J Royal Statist Soc Ser B* 76:243–246.
- Imbens GW, Rubin DB. 2015. *Causal inference in statistics, social, and biomedical sciences*. New York: Cambridge University Press.
- Kioumourtzoglou MA, Spiegelman D, Szpiro AA, Sheppard L, Kaufman JD, Yanosky JD, et al. 2014. Exposure measurement error in $PM_{2.5}$ health effects studies: A pooled analysis of eight personal exposure validation studies. *Environ Health* 13:2.
- Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. 2011. Assessing temporally and spatially resolved $PM_{2.5}$ exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos Environ* 45:6267–6275.
- Krewski D, Burnett RT, Goldberg MS, Hoover BK, Siemiatycki J, Jerrett M, et al. 2003. Overview of the reanalysis of the Harvard Six Cities Study and American Cancer Society Study of Particulate Air Pollution and Mortality. *J Toxicol Environ Health A* 66:1507–1551.
- 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.
- 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* 28:627–634, PMID: 28768298.
- Mathur MB, Ding P, Riddell CA, VanderWeele TJ. 2018. Web site and R package for computing E-values. *Epidemiology* 29:e45–e47.
- McCaffrey DF, Lockwood JR, Setodji CM. 2013. Inverse probability weighing with error-prone covariates. *Biometrika* 100:671–680.
- Nethery RC, Mealli F, Sacks JD, Dominici F. 2019. Causal inference and machine learning approaches for evaluation of the health impacts of large-scale air quality regulations. Available: <https://arxiv.org/abs/1909.09611> [accessed 27 June 2020].
- 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.
- Papadogeorgou G, Dominici F. 2020. A causal exposure response function with local adjustment for confounding: Estimating health effects of exposure to low levels of ambient fine particulate matter. *Ann Appl Stat* 14:850–871.
- Peng R. 2015. The reproducibility crisis in science: A statistical counterattack. *Significance* 12:30–32.
- Politis DN. 2003. The impact of bootstrap methods on time series analysis. *StatSci* 18: 219–230.
- Requia WJ, Di Q, Silvern R, Kelly JT, Koutrakis P, Mickley LJ, et al. 2020. An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. *Environ Sci Technol* 54:11037–11047.
- Robins JM, Hernan MA, Brumback B. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550–560.
- Rosenbaum PR. 2002. Overt bias in observational studies. In: *Observational Studies*. Springer Series in Statistics. New York: Springer.
- Rosenbaum PR, Rubin DB. 1983. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J Royal Stat Soc: Ser B* 45:212–218.
- Rubin DB. 2008. For objective causal inference, design trumps analysis. *Ann Appl Stat* 2:808–840.
- Rubin DB, Thomas N. 1996. Matching using estimated propensity scores: Relating theory to practice. *Biometrics* 52:249–264.
- Sarnat SE, Klein M, Sarnat JA, Flanders WD, Waller LA, Mulholland JA, et al. 2010. An examination of exposure measurement error from air pollutant spatial variability in time-series studies. *J Expo Sci Environ Epidemiol* 20:135–146.
- 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.
- Stanek LW, Brown JS, Stanek J, Gift J, Costa DL. 2011. Air pollution toxicology—a brief review of the role of the science in shaping the current understanding of air pollution health risks. *Toxicol Sci* 120 Suppl 1:S8–S27.

- Steiner PM, Cook TD, Shadish WR. 2011. On the importance of reliable covariate measurement in selection bias adjustments using propensity scores. *J Educ Behav Stat* 36:213–236.
- Stodden V, McNutt M, Bailey DH, Deelman E, Gil Y, Hanson B, et al. 2016. Enhancing reproducibility for computational methods. *Science* 354:1240.
- Stuart EA. 2010. Matching methods for causal inference: A review and a look forward. *Stat Sci* 25:1–21.
- Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. 2005. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *AmJ Epidemiol* 162:279–289.
- Szpiro AA, Sheppard L, Lumley T. 2011. Efficient measurement error correction with spatially misaligned data. *Biostatistics* 12:610–623.
- 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.
- U.S. Environmental Protection Agency. 2019. Chartered Clean Air Scientific Advisory Committee (CASAC) public teleconference on particulate matter (PM). Available: <https://yosemite.epa.gov/sab/sabproduct.nsf/MeetingCalCASAC/4F40665AD1DDCEF6852583A000645464?OpenDocument> [accessed 13 April 2020].
- U.S. Environmental Protection Agency. 2021. NAAQS table. Available: www.epa.gov/criteria-air-pollutants/naaqs-table [accessed 12 September 2020].
- U.S. Environmental Protection Agency. 2020a. Regulatory impact analyses for air pollution regulations. Available: www.epa.gov/economic-and-cost-analysis-air-pollution-regulations/regulatory-impact-analyses-air-pollution [accessed 12 September 2020].
- U.S. Environmental Protection Agency. 2020b. Integrated science assessment (ISA) for particulate matter. Available: www.epa.gov/isa/integrated-science-assessment-isa-particulate-matter [accessed 12 September 2020].
- Van der Laan MJ, Rose S. 2011. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer Series in Statistics. New York:Springer-Verlag.
- van Donkelaar A, Martin RV, Li C, Burnett RT. 2019. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. *Environ Sci Technol* 53:2595–2611.
- VanderWeele TJ, Ding P. 2017. Sensitivity analysis in observational research: Introducing the E-value. *Ann Intern Med* 167:268–274.
- Villeneuve PJ, Weichenthal SA, Crouse D, Miller AB, To T, Martin RV, et al. 2015. Long-term exposure to fine particulate matter air pollution and mortality among Canadian women. *Epidemiology* 26:536–545.
- Vodonos A, Awad YA, Schwartz J. 2018. The concentration-response between long-term PM_{2.5} exposure and mortality. A meta-regression approach. *Environ Res* 166:677–689.
- Webb-Vargas Y, Rudolph KE, Lenis D, Murakami P, Stuart EA. 2017. An imputation-based solution to using mismeasured covariates in propensity score analysis. *Stat Methods Med Res* 26:1824–1837.
- World Health Organization. 2005. WHO air quality standards. Available: [www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](http://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health) [accessed 27 June 2020].
- 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. *Ann Appl Stat* 13:520–547.
- Wu X, Braun D, Schwartz J, Kioumourtzoglou MA, Dominici F. 2020. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Sci Adv* 6:eaba5692.
- Wu X, Mealli F, Kioumourtzoglou MA, Dominici F, Braun D. In review. Matching on generalized propensity scores with continuous exposures. Available: <https://arxiv.org/abs/1812.06575> [accessed 12 September 2020].
- Zeiler M. 1999. Modeling our world: The ESRI guide to geodatabase design. Redlands, CA:ESRI, Inc.
- Zhu Y, Coffman DL, Ghosh D. 2015. A boosting algorithm for estimating generalized propensity scores with continuous treatments. *J Causal Interference* 3:25–40.
- 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.

HEI QUALITY ASSURANCE STATEMENT

The conduct of this study was subjected to independent audits by RTI International staff members Dr. Linda Brown and Dr. Prakash Doraiswamy. These staff members are experienced in quality assurance (QA) oversight for air quality monitoring, chemical transport modeling, use of satellite data, and epidemiological methods and analysis. The RTI QA oversight team also included statistician Dr. Sahar Zangeneh who reviewed the statistical methods and accompanying codes.

The QA oversight program consisted of an initial onsite audit of the research study at Harvard University for conformance to the study protocol and standard operating

procedures and a final remote audit of the final report and the data processing steps. The onsite audit was performed by Drs. Brown and Doraiswamy. The final remote audit was performed by Drs. Brown, Doraiswamy, and Zangeneh. The dates of the audits and reviews are listed below.

Audit 1: Onsite Audit at Harvard University, Boston, Massachusetts, April 25–26, 2018

The audit reviewed the following study components: progress reports; personnel and staff; adequacy of equipment and facilities; internal quality assurance procedures; air quality data processing and documentation; health data processing and quality checks; and backup procedures. Program codes were inspected to verify proper documentation. The codebook for the air pollution data was examined. The audit included an observation of the demonstration of selected script executions for the exposure estimates. No errors were noted, but recommendations were made for updating the study plan, expanding the quality plan, double-checking some model results that show unexpected results, documenting codes, documenting procedures and assumptions related to model development and QA/QC, developing a data dictionary/code book for the health data, and implementing QA/QC procedures for the health data to ensure independent checking of SAS codes for data management and analysis. The audit was conducted at an external location near Harvard and, therefore, did not include an inspection of facilities or equipment.

Audit 2: Final Remote Audit, October–December 2021

The final remote audit consisted of two parts: (1) review of the final project report, and (2) audit of data processing steps. The review of the final report focused on ensuring that the methods are well documented and the report is easy to understand. The review also examined if the report highlighted key study findings and limitations. In addition, this review provided guidance on specific aspects of the data processing sequence that could be reviewed remotely.

The data audit included (1) a remote live demonstration of selected data processing codes, and (2) the review of the codes for data reduction, processing and analysis, and model development. This specific portion of the audit was restricted to the key components of the study and associated findings. Selected codes (in R) for statistical model development were made available on GitHub. No data were sent to RTI due to data confidentiality restrictions. Therefore, data inputs to the codes were not available.

The codes were reviewed at RTI to verify, to the extent feasible, linkages between the various scripts, confirmation of the models reported, and verification of key tables. The codes appear to be consistent with the models and tables described in the report and followed the overall model development procedure described. The values themselves could not be generated at RTI due to unavailability of the input data. The

exposure datasets for $PM_{2.5}$ and O_3 were available on the NASA Socioeconomic Data and Applications Center (SEDAC) data repository. Selected datasets were downloaded and visualized to confirm logical temporal and spatial trends and agreement with the figures in the report.

The remote live demonstration included a real-time execution of selected codes generating keytables and figures in the report. Values generated by the codes during the real-time demonstration matched the values in the report after unit conversion and rounding. No major quality-related issues were identified from the review of the codes and the report. Minor recommendations were made for improved clarity and data accessibility.

A written report was provided to HEI. The QA oversight audit demonstrated that the study was conducted according to the study protocol. The final report, except as noted in the comments, appears to be representative of the study conducted.



Linda Morris Brown, MPH, DrPH, Epidemiologist, Quality Assurance Auditor



Prakash Doraiswamy, PhD, Air Quality Specialist, Quality Assurance Auditor



Sahar Zangeneh, PhD, Statistician, Quality Assurance Auditor
December 20, 2021

MATERIALS AVAILABLE ON THE HEI WEBSITE

Appendix A contains supplemental tables and figures not included in the main report. They are available on the HEI website at www.healtheffects.org/publications:

Figure 1. Estimated Causal ERC Relating $PM_{2.5}$ to All-Cause Mortality in Medicare Beneficiaries (2000–2016)

Table 1. HRs and 95% CIs Relating $PM_{2.5}$, NO_2 , and O_3 to All-Cause Mortality in Medicare Beneficiaries (2000–2016)

Table 2. Sensitivity Analysis Including Point Estimates and 95% CIs of the HRs Using Penalized Splines of Each Measured Potential Confounder

Figure 2. Pipeline for Processing Data from the CDC's BRFSS

Figure 3. Pipeline for Processing Data from the U.S. Census and American Community Survey

Figure 4. Pipeline for Processing Data from the Google Earth Engine

Figure 5. Pipeline for Cleaning and Merging Exposure, Covariate, and Health Data to Produce Combined Datasets Covering the Period 1999–2016

ABOUT THE AUTHORS

Francesca Dominici, Ph.D., is a co-director of the Data Science Initiative at Harvard University in Cambridge, MA, and the Clarence James Gamble Professor of Biostatistics, Population and Data Science at the Harvard T.H. Chan School of Public Health in Boston, MA. She is an elected member of the National Academy of Medicine and the International Society of Mathematical Statistics. She leads an interdisciplinary group of scientists with the ultimate goal of addressing important questions in environmental health science, climate change, and health policy. Her productivity and contributions to the field have been remarkable. She has provided the scientific community and policy makers with robust evidence on the adverse health effects of air pollution, noise pollution, and climate change. Her studies have directly and regularly affected air quality policy. She has published more than 220 peer-reviewed publications and was recognized in Thomson Reuters' 2019 list of the most highly cited researchers, ranking in the top 1% of cited scientists in her field. Her work has been covered by *The New York Times*, *The Los Angeles Times*, the BBC, *The Guardian*, CNN, and NPR. In April 2020, she was awarded the Karl E. Peace Award for Outstanding Statistical Contributions for the Betterment of Society by the American Statistical Association. She is an advocate for the career advancement of women faculty. Her work on the Johns Hopkins University Committee on the Status of Women earned her the campus Diversity Recognition Award in 2009. At the Harvard T.H. Chan School of Public Health, she has led the Committee for the Advancement of Women Faculty.

Antonella Zanobetti, Ph.D., is a principal research scientist in the Department of Environmental Health at the Harvard T.H. Chan School of Public Health. She has more than 20 years of experience in environmental epidemiology and with the Medicare cohort. Her research interests include studying the health effects of air pollution, temperature extremes, and climate change on mortality and hospital admissions, focusing on cardiovascular, respiratory, and neurological disorders endpoints. She is also interested in socioeconomic influences on health and environmental health disparities and developing innovative statistical methodologies to examine emerging issues in environmental epidemiology.

Joel Schwartz, Ph.D., is a professor of environmental epidemiology in the Department of Environmental Health at the

Harvard T.H. Chan School of Public Health in Boston, MA. 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. A world-renowned 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 of his 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 and deaths. This work has led to a tightening of U.S. air quality standards. He has done considerable work on the health effects of O₃ exposure. He has several international collaborations underway in this area. His recent work has focused on the cardiovascular effects of air pollution and on factors that modify responses to air pollution, suggesting that individuals with diabetes are more susceptible. He is also an expert in methods for causal inference and regression spline models, nonparametric smoothing, and generalized additive models. He has extensive expertise in the use of cost–benefit analysis to make environmental decisions. He has developed methodologies for assessing the benefits of lead control and applied those methodologies to the decision to remove lead from gasoline. Recently, in collaboration with colleagues at the Centers for Disease Control and Prevention, his work led to a decision to revise the Centers' lead screening recommendations for children. He is also involved in cost–benefit analysis of air pollution control.

Danielle Braun, Ph.D., is a senior research scientist working jointly at the Harvard T.H. Chan School of Public Health and the Department of Data Sciences at the Dana-Farber Cancer Institute, both in Boston, MA. 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, clinical decision support tools, genetic epidemiology, survival analysis, and frailty models. She has mentored undergraduate students, master's degree students, and Ph.D. candidates for more than eight years and co-leads the BayesMendel Lab in Boston, MA, in addition to leading many recent projects and working closely with Ph.D. candidates on their theses.

Ben Sabath, MA, is a data scientist whose work supports the research of the Dominici lab in Cambridge, MA, and its collaborators. His work involves developing software packages to aid the work of the lab on Harvard's high-performance computing system as well as preparing data for efficient analysis for researchers as part of the lab's data science platform.

Xiao Wu, Ph.D., was a postdoctoral research fellow in the Department of Biostatistics at the Harvard T.H. Chan School of Public Health in Boston, MA. He is currently a data science postdoctoral fellow at Stanford University in Stanford, CA. His research interests lie in developing statistical and causal inference methods to address methodological needs in climate and health research. His dissertation focuses on developing robust and interpretable causal inference methods to handle error-prone, continuous, and time-series exposures. He is also working on collaborative projects to design Bayesian clinical trials, meta-analyses, and real-world evidence studies.

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

- Braun D, Gorfine M, Parmigiani G, Arvold N, Dominici F, Zigler C. 2017. Propensity scores with misclassified treatment assignment: a likelihood-based adjustment. *Biostatistics* 18:695–710.
- Carone M, Dominici F, Sheppard L. 2019. In pursuit of evidence in air pollution epidemiology: the role of causally driven data science. *Epidemiology* 31:1–6.
- Cutler D, Dominici F. 2018. A breath of bad air: Trump environmental agenda may lead to 80 000 extra deaths per decade. *JAMA* 319:2261–2262.
- 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* 318:2446–2456.
- 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.
- Dominici F, Zigler CM. 2017. Best practices for gauging evidence of causality in air pollution epidemiology. *Am J Epidemiol* 186:1303–1309.
- Goldman G, Dominici F. 2019. Don't abandon evidence and process on air pollution policy. *Science* 363:1398–1400.
- Lee M, Schwartz J, Wang Y, Dominici F, Zanobetti A. 2019. Long-term effect of fine particulate matter on hospitalization with dementia. *Environ Pollut* 254(Pt A):112926.
- Lee K, Small D, Dominici F. 2018. Discovering effect modification and randomization in air pollution studies. Available: <https://arxiv.org/pdf/1802.06710.pdf> [accessed 27 June 2020].
- Makar M, Antonelli JL, Di Q, Cutler D, Schwartz J, Dominici F. 2017. Estimating the causal effect of lowering particulate matter levels below the United States standards on hospitalization and death. *Epidemiology* 28:627–634.
- Nethery R, Mealli F, Sacks J, Dominici F. 2019. Causal inference and machine learning approaches for evaluation of the health impacts of large-scale air quality regulations. Available: <https://arxiv.org/abs/1909.09611> [accessed 27 June 2020].
- Papadogeorgou G, Dominici F. 2020. A causal exposure response function with local adjustment for confounding: estimating health effects of exposure to low levels of ambient fine particulate matter. *Ann Appl Stat* 14:850–871.
- Rhee J, Dominici F, Zanobetti A, Schwartz J, Wang Y, Di Q, et al. 2019. Impact of long-term exposures to ambient PM_{2.5} and ozone on acute respiratory distress syndrome (ARDS) risk for older adults in the United States. *CHEST* 156:71–79.
- Schwartz JD, Yang W, Yitshak-Sade M, Dominici F, Zanobetti A. 2018. Estimating the effects of PM_{2.5} on life expectancy using causal modeling methods. *Environ Health Perspect* 126:127002.
- Thomas EG, Trippa L, Parmigiani G, Dominici F. 2020. Estimating the effects of fine particulate matter on 432 cardiovascular diseases using multi-outcome regression with tree-structured shrinkage. *J Am Stat Assoc*; 10.1080/01621459.2020.1722134
- Wei Y, Wang Y, Di Q, Choirat C, Wang Y, Koutrakis P, et al. 2019. Short term exposure to fine particulate matter and hospital admission risks and costs in the Medicare population: time stratified, case crossover study. *BMJ* 367:l6258.
- Wu X, Braun D, Schwartz J, Kioumourtoglou M, Dominici F. 2020. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Science Adv*; 10.1126/sciadv.aba5692.
- Wu X, Braun D, Kioumourtoglou MA, Choirat C, Di Q, Dominici F. 2019. Causal inference in the context of an error prone exposure: air pollution and mortality. *Ann Appl Stat* 13:520–547.
- Wu X, Mealli F, Kioumourtoglou MA, Dominici F, Braun D. In review. Matching on generalized propensity scores with continuous exposures. Available: <https://arxiv.org/pdf/1812.06575.pdf> [accessed 27 June 2020].
- Zigler CM, Choirat C, Dominici F. 2017. Impact of National Ambient Air Quality Standards nonattainment designations on particulate pollution and health. *Epidemiology* 29:165–174.

Research Report 211, *Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Implementation of Causal Inference Methods*, F. Dominici et al.

INTRODUCTION

Ambient air pollution is a significant contributor to the global burden of disease (GBD 2020; HEI 2020). Although air pollution concentrations have been declining over the past few decades in many higher-income countries, several studies published in the past decade have reported associations between risk of mortality and long-term exposures to fine particulate matter (PM_{2.5}*) even at low concentrations (e.g., Beelen et al. 2014a,b; Crouse et al. 2012, 2015; Hales et al. 2012; Pinault et al. 2016). To inform future risk assessment and regulation, it is important to confirm whether associations with mortality and other adverse health effects continue to be observed as air pollution concentrations decline still further. It is also important to better understand the shape of the exposure–response (ER) function at low concentrations. Both issues remain as major uncertainties for setting air quality standards in North America and Europe. The growing body of evidence demonstrating health effects at concentrations below current air quality standards, the large overall contributions of air pollution to the global burden of disease, and the general interest in reducing greenhouse gas emissions suggest that more stringent air quality standards and guidelines will likely be considered in the future.

As described in detail in the Preface to this Report, in 2016 HEI funded three studies under Request for Applications (RFA) 14-3 to explore the issue of health effects associated with exposures to low concentrations of air pollution using large cohorts and administrative databases. Dr. Dominici’s resulting study, *Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Implementation of Causal Inference Methods*, focused on a Medicare cohort in the United States. Additional information about

the RFA and the two other studies funded by HEI that were conducted in Canada and Europe is included in the Preface. It should be noted that all three study teams are conducting additional analyses to harmonize their approaches. Through this collaboration, the teams aim to (1) formally evaluate dose–response thresholds, (2) share analytical techniques and identify common statistical methods (e.g., a common set of covariates across the studies), and (3) determine strengths, weaknesses, and common findings of the three studies. That work is expected to be completed at the end of 2021.

Dominici’s study was conducted in two phases. In November 2019, HEI Published *Research Report 200: Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1*, along with an associated Commentary (Dominici et al. 2019). That Report and Commentary summarized and discussed analyses and findings produced through the first half of Dominici’s study. The present Commentary focuses on the research and findings produced during the second half of the study, recognizing that the work builds on the Phase 1 analyses.

This Commentary was prepared by the HEI Low-Exposure Epidemiology Studies Review Panel, which was convened to review these three HEI-funded studies, and members of the HEI Scientific Staff. The Commentary includes the scientific and regulatory background for the research, a summary of the study’s approach and key results, and the Panel’s evaluation of the Investigators’ Report (IR) highlighting strengths and weaknesses of the study. This Commentary is intended to aid the sponsors of HEI and the public by placing the IR 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, the European Union Ambient Air Quality Directives, and similar measures 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 public

Dr. Francesca Dominici’s 4-year study, “Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution,” began in March 2016. Total expenditures were \$2,427,228. The draft Investigators’ Report from Dominici and colleagues was received for review in January 2021. A revised report, received in June 2021, was accepted for publication in July 2021. During the review process, the HEI 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 Panel’s Commentary.

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

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

health effects likely to be seen at various levels of the standards;

- Identifying and setting standards based on risk assessments;
- Monitoring air quality to identify areas that do not meet the standards; and
- Implementing air quality control interventions to meet the standards by reducing the concentrations to which people are exposed.

SETTING AIR QUALITY STANDARDS IN THE UNITED STATES

The U.S. Clean Air Act requires that in setting the National Ambient Air Quality Standards (NAAQS), the U.S. Environmental Protection Agency (U.S. EPA) Administrator reviews all available science and sets the NAAQS for all major (criteria) pollutants (e.g., particulate matter [PM], nitrogen dioxide [NO₂], and ozone [O₃]) at a level “requisite to protect the public health with an adequate margin of safety.” In practice, that review has had two principal steps:

1. Synthesis and evaluation of all available science in what is now called an Integrated Science Assessment. This document reviews the widest range of exposure, dosimetry, toxicological, mechanistic, clinical, and epidemiological evidence. It then — using a predetermined set of criteria (U.S. EPA 2015) — draws on all lines of evidence to determine whether the exposure is causal, likely to be causal, or suggestive of being causal 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 concentrations at which health effects are observed, the likely implications of such concentrations 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.

The Risk and Policy Assessment also examines the uncertainties around estimates of health effects and the shape of the ER function, especially at concentrations near and below the then-current NAAQS. Although a range of possible shapes for the ER functions is considered, including whether there is a threshold at a concentration below which effects are not likely, the U.S. EPA's conclusions in these reviews thus far have not found evidence of such 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 under the Clean Air Act at “a level requisite to protect public health with an adequate margin of safety,” it has been understood that there are likely additional, albeit more uncertain, health effects of exposure to air pollution concentrations below the NAAQS.

Both documents are subjected to extensive public comments and review by the Clean Air Scientific Advisory Committee, which was established under the U.S. Clean Air Act. The Committee is charged with peer-reviewing the documents, which includes advising the Administrator on the strength and uncertainties in the science and making the decision whether to retain or change the NAAQS. The current NAAQS for longer-term exposure to PM_{2.5}, NO₂, and O₃ are as follows (<https://www.epa.gov/criteria-air-pollutants/naaqs-table>):

- PM_{2.5}: annual mean averaged over 3 years of 12 µg/m³;
- NO₂: annual mean of 53 ppb (approximately 100 µg/m³); and
- O₃: annual fourth-highest daily maximum 8-hour concentration, averaged over 3 years, of 70 ppb (approximately 140 µg/m³).

EVALUATING ASSOCIATIONS BELOW CURRENT AIR QUALITY STANDARDS AND GUIDELINES

As the quality and availability of data on air pollution concentrations improved over the first decade of this century, results from new studies began to emerge starting in 2012 (e.g., in Canada, Crouse et al. 2012; and, in New Zealand, Hales et al. 2012) that suggested that associations between PM and mortality could be observed down to concentrations well below the NAAQS of 12 µg/m³. For example, associations with mortality were present in the Canadian study at PM_{2.5} concentrations of only a few micrograms per cubic meter. These two studies found robust associations, with some evidence of even larger effects at the lowest concentrations of PM_{2.5}, but neither examined associations with exposures to NO₂ or O₃. If replicated in other populations and by other investigators, such findings 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, the findings suggested several questions:

- Would the results be robust to the application of a range of alternative analytic models and their uncertainties?
- Could other important determinants of population health — such as age, socioeconomic status (SES), health status, access to medical care, and differences in air pollution sources and time-activity patterns — modify or confound the associations seen?
- Would the results change if risk estimates were more fully 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?

As described in the Preface, these important questions were the basis for RFA 14-3. After a rigorous selection process, the Research Committee recommended the study by Dominici

and colleagues for funding because it thought the study had many strong aspects, such as the very large sample size, U.S.-wide coverage, and the experienced team. The development of numerous new methods (mainly causal modeling methods) that had the potential for wider use was also considered a strength.

SUMMARY OF APPROACH AND METHODS

The overarching purpose of the Dominici study was to address some of the knowledge gaps related to health effects of long-term exposures to low concentrations of air pollution. The study encompassed several goals related to modeling spatial and temporal patterns of ambient air pollution, developing causal inference statistical models, and describing risks to morbidity and mortality associated with exposures to pollution. The investigators presented results from conventional regression models and from the newly developed causal approaches. The analyses were conducted in a national-level administrative cohort of over 68 million older American adults. Throughout the study, the investigators examined health effects for the entire cohort and for a subpopulation exposed to annual average concentrations of $PM_{2.5}$ below or equal to $12 \mu\text{g}/\text{m}^3$ during every year of follow-up (henceforth referred to as the low-exposure cohort). Also underlying this study was an effort to make the methods and data available to the wider scientific community.

STUDY OBJECTIVES

The 4-year study had four broad aims, some of which were addressed in the first phase of the study (see Commentary Table). Here, the focus was on the analyses presented and discussed in the Final Report, as follows.

Aim 1. Exposure Prediction and Data Linkage Estimate long-term exposures to low concentrations of ambient $PM_{2.5}$ mass, O_3 , and NO_2 at high spatial resolution (1 km by 1 km) for the contiguous United States during the period 2000–2016, by applying and extending hybrid prediction models that use ground monitoring, land use, and meteorological data and satellite observations in conjunction with chemical-transport models. Link these predictions to health data while accounting for the misaligned nature of the data.

Aim 2. Causal Inference Methods for Exposure–Response Functions Develop a new causal inference framework that is robust to model misspecification for confounding and to account for exposure error. Specifically, develop new methods to estimate a nonlinear ER function while accounting for exposure error, adjust for measured and unmeasured confounders, and detect effect modification in the presence of multiple exposures.

Aim 3: Evidence of Adverse Health Effects Apply methods developed in Aim 2, along with traditional regression approaches, to estimate all-cause mortality by year and zip code associated with long-term exposure to ambient air pollution for U.S. Medicare enrollees 65 years of age or older between 2000 and 2016. Examine health effects for the entire cohort and the low-exposure cohort.

Aim 4: Tools for Data Access and Reproducibility Develop approaches for data sharing, record linkage, and statistical software so that other researchers can use the data and analytical methods to foster transparency and reproducibility of the work.

METHODS AND STUDY DESIGN

Exposure Modeling

This Final Report summarizes the development of predicted exposures to daily average $PM_{2.5}$ (Di et al. 2019), O_3 (Requia et al. 2020), and NO_2 (Di et al. 2020) at a 1-km \times 1-km grid for the contiguous United States during the period 2000–2016. For O_3 , daily maximum 8-hour ground-level concentrations were estimated for warm-weather months. Predictions were developed using a previously developed and validated ensemble model that uses multiple machine learning algorithms and predictor variables from multiple sources.

Model inputs included monitoring data from the U.S. EPA Air Quality System, satellite-derived aerosol optical depth, meteorological variables from the North American Regional Reanalysis data set, land-use variables that represent local emissions and small-scale variations in concentrations (e.g., road density, elevation, and normalized difference vegetation index), and daily predictions from two chemical-transport models—the global GEOS-Chem model and the regional-scale Community Multiscale Air Quality Model—to simulate atmospheric components. Dominici and colleagues applied a geographically weighted generalized additive model as an ensemble model that blended predicted concentrations from three types of machine learning models to predict air pollution concentrations. Missing data were imputed using machine learning and linear interpolation. In a final step, temporally and spatially lagged predictions from nearby monitoring sites and neighboring days were added to the model to predict air pollution concentrations.

The ensemble models for each pollutant were validated with 10-fold cross-validation. This method entails performing the fitting procedure 10 times, with each fit being performed on a training set consisting of 90% of the total monitoring data selected at random and the remaining 10% used as a hold-out set for validation. The investigators then aggregated the cross-validated results from the 10 runs and compared them with the corresponding monitoring values by site and

Commentary Table. Comparison of Study Accomplishments in Phase 1 (HEI Report 200) and Phase 2

Study Aims	Phase 1	Added in Phase 2
Aim 1: Exposure prediction	<ul style="list-style-type: none"> Summarized exposure predictions for daily $PM_{2.5}$ and O_3 at 1-km \times 1-km grid for the contiguous United States over the period 2000–2012, using an ensemble modeling approach (Di et al. 2019). 	<ul style="list-style-type: none"> Extended exposure predictions for $PM_{2.5}$ and O_3 to 2016. Summarized exposure predictions for NO_2 and O_3 using a similar approach for the contiguous United States over the period 2000–2016 (Di et al. 2020; Requia et al. 2020).
Aim 2: Causal inference methods	<ul style="list-style-type: none"> Developed a new statistical method for causal inference to reduce bias due to exposure measurement error and unmeasured confounding. The method combines a regression calibration-based adjustment for a continuous error-prone exposure with generalized propensity scores to adjust for potential confounding (Wu et al. 2019). Developed LERCA, a flexible new method for causal inference to estimate an ER function with local adjustment for confounding (Papadogeorgou and Dominici 2020). The method allows for variation in confounders and strength of confounding at various exposures, model uncertainty about confounder selection and the shape of the ER function, and assessment of the observed covariates' confounding importance at various exposures. 	<ul style="list-style-type: none"> Developed nonparametric causal inference methods that use a generalized propensity score matching, weighting, and adjustment to estimate the causal ER function for air pollution exposure on mortality. Exposure to air pollution was set as a continuous variable that is computationally tractable and scalable to handle large datasets.
Aim 3: Epidemiological studies	<ul style="list-style-type: none"> Conducted case-crossover study of short-term exposure to $PM_{2.5}$ and O_3 and all-cause mortality in Medicare enrollees 2000–2012, including effects among those in a low-exposure cohort, Medicaid-eligible group, and other subgroups (Di et al. 2017a). Conducted cohort study of long-term exposure to $PM_{2.5}$ and O_3 and all-cause mortality in Medicare enrollees 2000–2012, using the Anderson-Gill model of Cox regression. Sensitivity analysis of effects was conducted among those in a low-exposure cohort and Medicaid-eligible group (Di et al. 2017b). Analysed a Medicare Current Beneficiary Survey subsample, which has information on individual risk factors, to assess the sensitivity of results to omission of several individual-level confounders. 	<ul style="list-style-type: none"> Implemented five statistical approaches to estimate the effects of long-term $PM_{2.5}$ exposure on all-cause mortality in Medicare enrollees aged 65 years and older from 2000 to 2016, accounting for potential confounders. The methods used were two traditional approaches that rely on regression for confounder adjustment (Cox proportional hazards and Poisson) and three causal inference methods (described in Aim 2). Estimated the effect in the low-exposure cohort. Applied the new matching method to estimate the ER functions for long-term exposures to $PM_{2.5}$, NO_2, and O_3 on all-cause mortality in single-pollutant models and multipollutant models where each individual pollutant was adjusted for the other two. In sensitivity analyses, estimated the HRs (under an assumption of a constant HR) for the three pollutants adjusted for the other two pollutants, using both the matching method and multivariate Poisson regression.
Aim 4: Data and methods availability	<ul style="list-style-type: none"> Provided code for implementation of new causal inference methods. 	<ul style="list-style-type: none"> Made exposure estimates for $PM_{2.5}$ available for public access. Documented data sources, analytical data sets, and statistical code to assist others who seek to reproduce the results.

LERCA = local exposure response confounding adjustment.

day to obtain the total R^2 , an indication of model fit. They regressed the difference between predicted and monitored $PM_{2.5}$ at a given site at a given time with the annual mean at the same site to derive a temporal R^2 (Kloog et al. 2011). They also compared the annual mean between monitored and predicted values at each site to derive a spatial R^2 .

Study Population

All analyses presented in the Final Report were based on adults 65 years of age or older who are beneficiaries of Medicare, the U.S. federal health insurance program for people who are 65 years of age or older or permanently disabled. Individuals enroll in Medicare upon reaching age 65 or incurring a qualifying disability and are followed until death. In Phase 2, the enrollment period was extended from 2000–2012 (as used in Phase 1) to 2000–2016, increasing the number of participants from 61 million to 68.5 million. Individual data obtained from the Centers for Medicare & Medicaid Services were 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. Medicaid is a program that provides health insurance coverage to low-income individuals; the investigators used Medicaid eligibility as a proxy variable to indicate low SES. Originally the investigators had planned also to investigate the health effects of low levels of air pollution in the Medicare Current Beneficiary Survey subsample, but those analyses were not included in the Final Report.

Exposure Assignment

Predicted annual average $PM_{2.5}$, NO_2 , and O_3 exposures were assigned to cohort participants' residential zip code for each year of follow-up. Zip codes vary in size based on population density and can cover a neighborhood in dense urban areas or represent an entire town, community, or area elsewhere. For example, zip codes are on average 24 km² in Los Angeles County, California, and 268 km² in the state of Texas. In total, there are about 42,000 zip codes in the United States, with a mean area of 234 km², comprising an average of 7,755 individuals per zip code.

For standard zip codes, Dominici and colleagues averaged the predicted daily pollutant concentrations for all 1-km² grid cells whose centroids fell within that zip code area. For zip codes that designated post office box locations, average concentrations were calculated by linking to the predictions from the nearest 1-km² grid cell. Annual averages were estimated by averaging the daily concentrations. Ultimately, they assigned the estimated annual zip code-level average pollutant concentration to all individuals who lived in that zip code for each calendar year. In this way, all cohort members were assigned time-varying, annual estimates of exposures to all three pollutants for every year of follow-up.

Main Epidemiological Analyses

For the main analysis in Phase 2, the investigators reanalyzed the effect of annual $PM_{2.5}$ exposure on all-cause mortality in the Medicare cohort with follow-up from 2000 to 2016, expanding their analytical methods to include a computationally efficient Poisson regression model and three causal inference approaches (matching, weighting, and adjustment) in addition to the Cox regression method used in Phase 1. These five approaches are summarized below.

The unit of analysis for most of the data used in Phase 2 (specifically, the Poisson and three causal inference approaches) was at the zip code level each year. By using estimates of exposure at the aggregated (i.e., zip code) level, along with similarly aggregated covariate values for many potential confounders, the analyses introduce aspects of an ecological study design to the analysis of a large cohort of individuals. This was highlighted by the authors' presentation of the equivalence of the Cox and Poisson models, and all approaches except for the Cox model were explicitly fitted to aggregated data. The use of aggregated exposures and potential confounders created a hybrid design that allowed for some individual-level covariates and adjustments in an otherwise area-aggregated analysis. The hybrid design introduced important statistical questions about the potential effect of measurement error and confounding on the results.

Cox Proportional Hazards Survival Model (Anderson-Gill Variant) As in Phase 1, the investigators used Cox proportional hazards models with individual-level data, stratified by selected individual-level covariates available from the Medicare database (i.e., 5-year age band, race and ethnicity, sex, and Medicaid eligibility). The data were adjusted for zip code- and county-level indicators for smoking behavior, body mass index, SES, race, education, and population density from the U.S. Census, the American Community Survey, and the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System. To account for potential residual or unmeasured spatial and temporal confounding, models were also adjusted for zip code-level meteorological variables, an indicator of broad geographic region (West, Midwest, South, and Northeast), and calendar year. Annual average concentrations of $PM_{2.5}$, NO_2 , or O_3 were the time-varying exposures, and likelihood of survival in a given follow-up year was the outcome.

Poisson Regression The Poisson regression modeling approach used annual predicted $PM_{2.5}$ as the time-varying exposure and the count of deaths at the given follow-up year, calendar year, and zip code as the outcome. To adjust for potential confounding, the Poisson model included the same zip code- or county-level time-varying covariates, region indicator variable, and calendar year variable as were included in

the Cox models. Also, as in the Cox model, strata-specific baseline risk rates were accounted for by stratifying on individual-level characteristics from the Medicare data.

Causal Inference Approaches The causal inference methods introduced in Phase 2 use generalized propensity scores. This approach attempts to mimic a study in which participants are randomly assigned to an exposed group and a reference group, such that potential confounders that are known to affect participants' mortality (such as sex, age, and Medicaid eligibility) can be assumed to be balanced between the two groups. Propensity scores were estimated by modeling the zip code-level exposure conditional on area-level risk factors, meteorological variables, and year and region, using gradient boosting (Chen and Guestrin 2016; Zhu et al. 2015). Thus, unlike in most causal inference analyses that estimate propensity scores for individuals, the investigators estimated propensity scores at the zip code level, thereby seeking covariate balance at this level.

Propensity score methods typically assume a dichotomous exposure (i.e., an exposed versus a less exposed or unexposed reference population). Therefore, the investigators developed and implemented novel generalized propensity score approaches to accommodate the continuous air pollution exposures in the study.

Three different causal modeling approaches using generalized propensity scores — matching, weighting, and adjustment — were applied to create an artificial population in which the covariate distributions did not differ by exposure status. This is important as covariate balance indicates the effectiveness of the causal inference approach at mimicking a randomized experiment based on known factors and thus informs the degree to which one can make a valid causal assessment. The three approaches for balancing the covariates in the data based on propensity scores are summarized below.

Matching The objective of matching is to construct datasets that approximate a randomized experiment as closely as possible by pairing exposed and unexposed observations to achieve good covariate balance. In a continuous exposure setting, the challenge is that it is unlikely that two units will have the exact same exposure; thus how to match becomes substantially more complicated. To overcome this challenge, the researchers developed a new matching approach (described in Wu et al. in review) to achieve covariate balance in a continuous exposure setting. The new method uses a nearest-neighbor caliper that matches zip codes on both the estimated scores and exposures. The closeness of exposure guarantees that the matched unit is a valid representation of observations for a particular exposure, and the closeness of the propensity scores ensures that there is proper adjustment for confounding. A Poisson regression model was then fitted on the matched dataset regressing the death count on $PM_{2.5}$ exposure, with person-time as the offset term, and stratifying

by the four individual-level characteristics and the follow-up year.

Weighting In this approach, following Robins and colleagues (2000), the inverse of the estimated generalized propensity scores was used to weight each observation and achieve covariate balance in an artificial, or pseudo, population. A weighted Poisson regression model was then fit for the pseudo population, regressing the death count on $PM_{2.5}$ exposure, with person-time as the offset term, incorporating the assigned weights, and stratifying by the four individual-level characteristics and the follow-up year.

Adjustment Following Hirano and Imbens (2004), the investigators included the estimated generalized propensity scores in the outcome model as a covariate. The conditional expectation of death counts, given the exposure and the estimated propensity scores, was modeled as a Poisson regression stratified by age, race, sex, Medicaid eligibility, and follow up-year, with an offset for person-year. In this approach, unlike with the matching and weighting approaches where the analysis is complete after fitting the Poisson regression model, the coefficients from the Poisson regression model do not provide causal interpretation; rather, the causal outcome analysis is conducted on the counterfactuals predicted by the Poisson model.

Additional Epidemiological Analyses

Low-Exposure Analyses In addition to the full cohort, Dominici and colleagues also performed the five statistical analyses on the sub-cohort of Medicare enrollees who were exposed to $PM_{2.5}$ concentrations lower than $12 \mu\text{g}/\text{m}^3$ during every year of follow-up (i.e., the low-exposure cohort described earlier). This exposure cut point was selected because $12 \mu\text{g}/\text{m}^3$ is the current NAAQS for long-term exposure to $PM_{2.5}$.

Lastly, Dominici and colleagues applied the newly developed generalized propensity scores matching method to estimate ER functions for all-cause mortality and long-term exposure to $PM_{2.5}$, NO_2 , and O_3 both individually and for each pollutant adjusted by the other two (Wu et al. 2019). The highest and lowest 1% of pollutant exposures were excluded to avoid instability at the boundaries.

Sensitivity Analyses The investigators presented several sensitivity analyses. For example, the cohort data for 2000–2012 were reanalyzed to assess how results changed with exposure data updated through 2016. Additionally, the analysis was repeated without year as a covariate to evaluate model sensitivity to unmeasured confounders that vary over time.

SUMMARY OF KEY FINDINGS

MODELING AND EXPOSURE ESTIMATION RESULTS

Dominici and colleagues reported good model performance, with a 10-fold cross-validation R^2 of 0.86 for daily $PM_{2.5}$ exposure predictions and lower exposure error at low concentrations. Results for NO_2 exposure predictions indicated good model performance, with a 10-fold cross-validation R^2 of 0.79 overall, a spatial R^2 of 0.84, and a temporal R^2 of 0.73, and good performance outside of metropolitan areas and in rural areas. For O_3 predictions, they obtained a 10-fold cross-validation R^2 of 0.90, a spatial R^2 of 0.86, and a temporal R^2 of 0.92, indicating good model performance, with better performance in the East North Central region and during summer. The mean estimate of $PM_{2.5}$ as assigned to cohort participants was 9.8 $\mu\text{g}/\text{m}^3$ (standard deviation 3.2). The report did not provide descriptive information about exposure estimates for the other pollutants.

EPIDEMIOLOGICAL RESULTS

In the main analysis, Dominici and colleagues found consistent, statistically significant results across their five statistical approaches. HRs and 95% confidence intervals (CIs) associated with a 10- $\mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ exposure were 1.07 (1.06, 1.07) for the traditional Cox regression, 1.06 (1.06, 1.07) for the traditional Poisson regression, 1.07 (1.05, 1.08) for the Poisson with general propensity score matching, 1.08 (1.07, 1.09) for the Poisson with inverse propensity score weighting, and 1.07 (1.06, 1.08) for the Poisson with propensity score adjustment (see Commentary Figure 1 and IR Table 2). Covariate balance for each model was evaluated using mean absolute correlation, with values <0.1 indicating successful randomization. The investigators showed that this value was smaller than 0.1 using their propensity score matching and weighting approaches (IR Figure 5).

Across all models, the investigators found notably larger effect estimates for the low-exposure cohort. For example, in the standard Cox models, they reported a HR of 1.37 (95% CI, 1.34 to 1.40) in the low-exposure cohort compared with a HR of 1.07 (95% CI, 1.06 to 1.07) in the full cohort (Commentary Figure 1 and IR Table 2). A recent review of studies that investigated associations between natural-cause mortality and $PM_{2.5}$ reported a greater relative risk in a meta-analysis of studies conducted at mean annual concentrations below 10 $\mu\text{g}/\text{m}^3$ than among all studies and among those conducted at mean concentrations below 25 $\mu\text{g}/\text{m}^3$ (Chen and Hoek, 2020). The larger effects reported in the low-exposure groups could also be due in part to those in the low-exposure group being more susceptible to the effects of exposure. For example, the low-exposure cohort excluded participants in large areas of the Eastern United States and likely excluded most people in New York, Los Angeles, and most major cities. That is to

say, the main analyses to some extent describe the risk for the elderly U.S. population as a whole, while the low-exposure analyses to some extent describe the risk for those in smaller towns and rural areas (who tend to be of lower SES, have lower levels of educational attainment, have poorer health behaviors, have poorer access to health services, and have a higher prevalence of diabetes or other comorbidities that might also increase susceptibility to the effects of exposure [Coughlin et al. 2019; O'Neill et al. 2003]).

Notably, however, at exposure levels below or equal to 12 $\mu\text{g}/\text{m}^3$, the causal inference approaches produced smaller estimates of the HRs than the traditional regression approaches suggesting that some of the enhanced risk may be due to confounding and/or model misspecification.

When restricted to the 2000–2012 population (as described in Phase 1), results were consistent with the 2000–2016 results. When year was excluded as a covariate, the estimated HRs were larger in magnitude, a possible indication of bias due to confounding by time, which was not addressed in the Phase 1 report and thus flagged in that Commentary. It is interesting to note that although confounding by time did inflate the associations, important positive associations remained between mortality and $PM_{2.5}$ after adjustment for time trends in Phase 2 of the project.

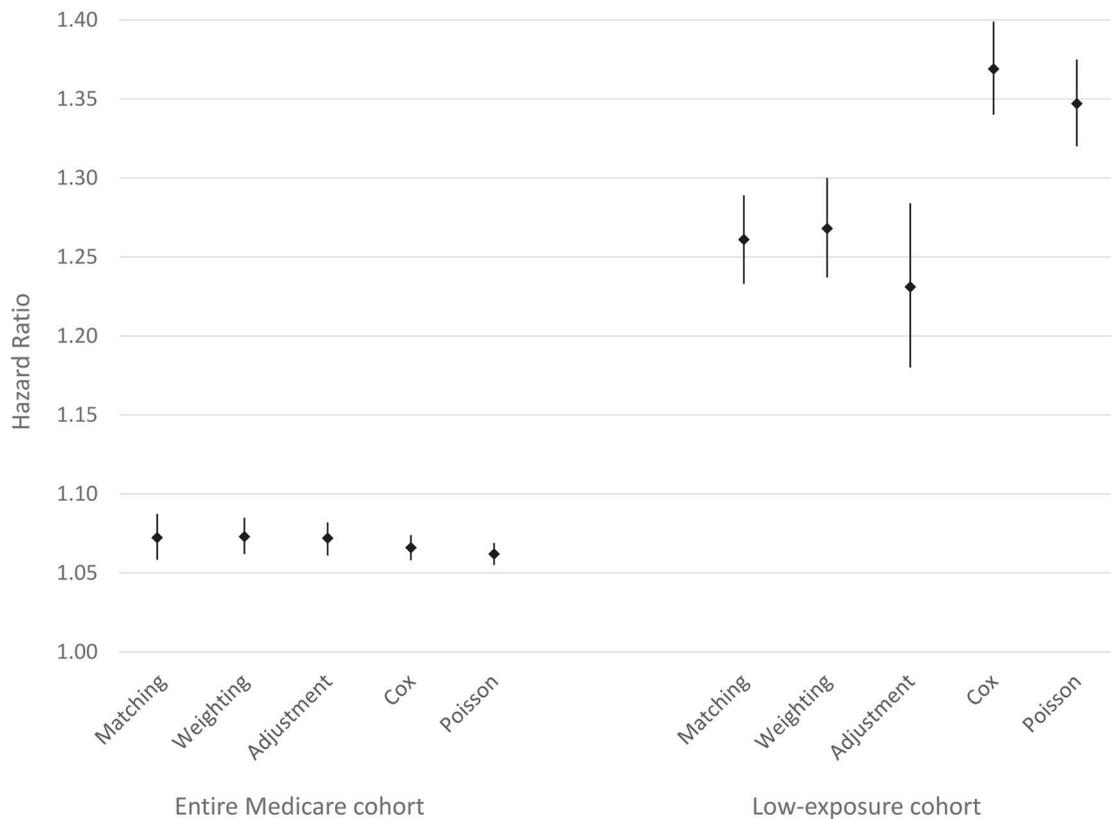
Exposure–Response Functions

Commentary Figure 2 summarizes the ER functions associated with long-term exposure to $PM_{2.5}$, NO_2 , and O_3 and all-cause mortality in the Medicare population 2000–2016, using HRs from a generalized propensity score matching analysis. In the single-pollutant models, Dominici and colleagues found evidence of increased risk of mortality associated with long-term $PM_{2.5}$ exposures across the range of annual average $PM_{2.5}$ concentrations between 2.77 and 17.16 $\mu\text{g}/\text{m}^3$, which included 98% of observations. The ER functions for $PM_{2.5}$ were almost linear at exposures below current U.S. standards, indicating adverse effects even at these low exposures.

The investigators' propensity score matching analysis also found evidence of a relationship between mortality and long-term exposures to NO_2 at the higher exposure concentrations. Associations at exposures lower than annual mean ≤ 53 ppb, the equivalent of the current U.S. annual NAAQS, were non-linear and statistically uncertain.

Similarly, their ER functions derived from propensity score matching for long-term O_3 exposures and mortality showed some evidence of increased risks at exposures higher than 45 ppb. The ER function was, however, almost flat at concentrations below 45 ppb, showing no statistically significant effect.

Generally, adjusting for the other two pollutants in the causal inference approach slightly attenuated the effects of $PM_{2.5}$ on mortality and slightly elevated the effects of NO_2 exposure, while results for O_3 remained almost unchanged.



Commentary Figure 1. Associations between longer-term exposures to $PM_{2.5}$ and all-cause mortality among enrollees in the full Medicare cohort (left side) and in the low-exposure cohort (right side). Data shown are HRs and 95% CIs. The HRs were estimated under five statistical approaches: three causal inference approaches using generalized propensity scores (matching, weighting, and adjustment) and two traditional approaches (Cox and Poisson regression). The HRs were calculated per $10\text{-}\mu\text{g}/\text{m}^3$ increase in $PM_{2.5}$ exposure. Results are presented for fully adjusted models. (Source: Adapted from Figure 6 in the Investigators' Report.)

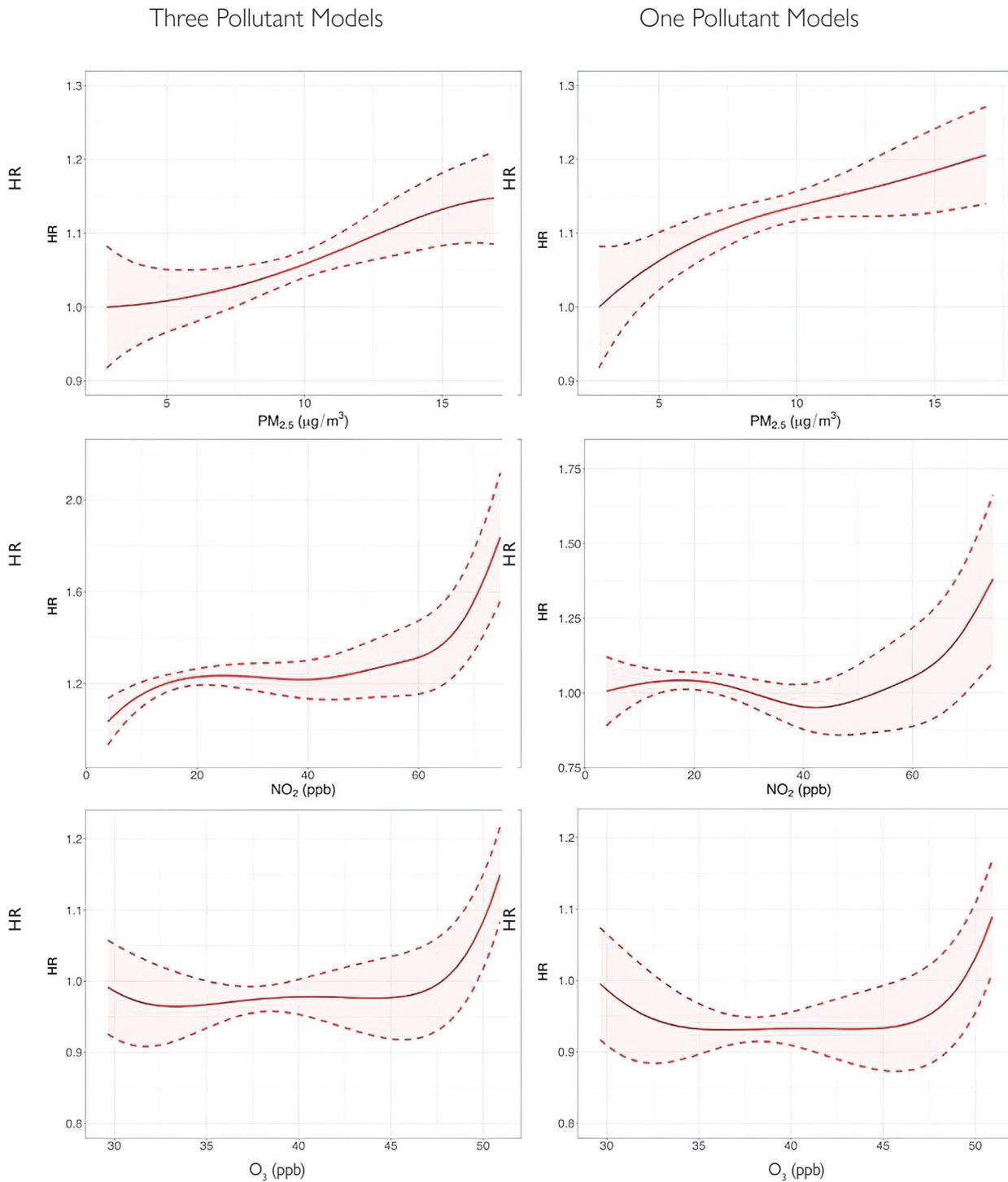
Reproducible Research

To allow for transparency and to support reproducibility of the research, the investigators were committed to sharing their data and statistical code. They have made the daily 1-km $PM_{2.5}$ predictions across the contiguous United States for years 2000–2016 available on a publicly accessible website in both RDS and GeoTiff formats at <https://beta.sedac.ciesin.columbia.edu/data/set/aqdh-pm2-5-concentrations-contiguous-us-1-km-2000-2016>.

In addition, they have posted their workflows and statistical codes for merging datasets and for running statistical analyses, along with most of their data, with the objective of developing an open science research data platform, at <https://github.com/NSAPH/National-Causal-Analysis>. Not all data can be made available, because of privacy restrictions (i.e., the Medicare data) or because the files were too large. In all cases, where the investigators were unable to share data directly, they have provided instructions on how to acquire and prepare the data for analyses.

EVALUATION BY THE HEI LOW-EXPOSURE EPIDEMIOLOGY STUDIES REVIEW PANEL

The HEI Low-Exposure Epidemiology Studies Review Panel concluded that this report presents a high-quality and thorough investigation into associations between risk of mortality and exposures to ambient air pollution in the United States. Importantly, the findings from the report contribute to our knowledge of effects on health associated with long-term exposures to low concentrations of ambient air pollution. In summary, Dominici and colleagues showed that the mean estimate of exposure to $PM_{2.5}$ among about 68 million Medicare cohort participants was just below $10\text{ }\mu\text{g}/\text{m}^3$. They reported consistent increases in risk of all-cause mortality ranging from 6% to 8% per $10\text{ }\mu\text{g}/\text{m}^3$ in $PM_{2.5}$ for five separate epidemiological approaches (see Commentary Figure 1) even after adjusting for key copollutants, providing strong evidence that mortality is associated with long-term exposures to $PM_{2.5}$. In the case of exposures to O_3 and NO_2 , although the investigators reported adverse associations with mortality, these were not found at the lowest concentrations.



Commentary Figure 2. Estimated ER functions relating $PM_{2.5}$, NO_2 , and O_3 to all-cause mortality among Medicare enrollees (2000–2016) with and without adjustment for copollutants. Data shown are HRs with 95% CIs obtained using a generalized propensity score matching approach. The left panels show the ER functions associating long-term exposure to one pollutant with all-cause mortality, adjusted for the other two pollutants as potential confounders. The right panels show the ER functions for single-pollutant models without adjusting for the other two pollutants. To avoid potentially unstable behavior at the support boundaries, the highest 1% and lowest 1% of pollutants exposures were excluded. (Source: Figure 7 in the Investigators’ Report.)

Particularly strong aspects of this work include the use of an extremely large, national health cohort (Medicare) with almost 70 million participants; relatively high-resolution annual mean exposure estimates for each year of follow-up; and the development of novel approaches to causal modeling to assess the associations between air pollution exposure and mortality. The development and presentation of five approaches to risk estimation was a major achievement of this work. The evaluation of the nonlinearity in multipollutant models was an additional valuable contribution. The Panel also appreciated that the datasets (those not subject to confidentiality restrictions) and statistical codes developed for the study have been made publicly available, thus facilitating transparency and reproducibility.

In spite of these many strengths, the Panel noted a few limitations with some of the approaches used, such as the quality of the exposure estimates in rural areas; the fact that all exposure estimates were aggregated to the zip code level of analysis; and the hybrid nature of the study design, which included some covariates measured at the individual level, others at the zip code level, and others at the county level. These and other aspects of the study design and approach and the interpretations of the findings and results are described and discussed in the following sections.

EVALUATION OF STUDY DESIGN AND APPROACH

Air Pollution Models and Exposure Estimation

The development of annual exposure estimates for three pollutants covering the contiguous United States was an impressive achievement of the study. This accomplishment is impressive because of the large geographic scope of the exposure models, because of the vast amount and variety of datasets the investigators assembled to produce them, and because of the computational requirements to do so. These exposure models allowed the investigators to assign exposure estimates to cohort participants, including those in rural areas where there are few or no pollution monitors, for each year of follow-up. The Panel had concerns, however, about the quality and accuracy of the estimates for rural areas, precisely because there are few or no pollution monitors. Generally, U.S. EPA monitors are located for the purpose of compliance with NAAQS, so they are placed in more populated, urban areas where air pollution concentrations are higher. Consequently, rural areas — where population densities and pollutant concentrations are lower — are not monitored as intensively. Thus, the models can be more prone to larger errors there, and they can't be validated as well as at other locations. Given that relatively few people live in these areas, the errors might not have much effect on the overall exposure estimates or the main epidemiological analyses. If these rural populations represent a sufficiently large portion of those with the lowest exposures, however, the errors introduced

here could be particularly important for the study in its influence on the low ends of the ER functions and on subsequent epidemiological analyses.

Generally, the Panel was impressed with the achievement of producing the models at the relatively fine spatial scale of 1 km by 1 km. However, models at this spatial resolution do not capture fine-scale variability in ambient concentrations; that is, they do not capture local gradients in concentrations, such as those along roadways or near major point sources. The exposure estimations for those living in the vicinity of such areas are therefore probably underestimated (for $PM_{2.5}$) or overestimated (for O_3 , because of local area scavenging).

Regardless, the investigators did not have access to full address information for cohort participants and therefore had to aggregate these pollution estimates to the geographic scale of zip codes for the purpose of estimating participants' long-term exposures. This analytic step entailed that all participants living in a given zip code, which in many cases can be 100–200 km² in size or more, were assigned the same exposure estimate. An implication of this fact is that the observed associations with mortality might be driven by larger-scale, pollution trends as opposed to highly localized gradients, such as might be found along roadways or near key point sources. As noted above, zip codes vary substantially in size, with rural zip codes generally covering much larger areas than urban zip codes. This might imply greater exposure error in rural areas, which might also have the lowest concentrations.

Ultimately, the methods for developing the models, and the models themselves, should prove valuable to other researchers who are studying air pollution and health, given that the exposure estimates have been made publicly available to access and download. From the inception of the study, the Panel was pleased to note that the investigators planned to make their data and methods available to other investigators. The Panel commends them for this effort to support research transparency and reproducibility, while also noting that not all of the datasets could be made available for free (for example, users must pay to access Medicare records from the U.S. Centers for Medicare & Medicaid Services' Research Data Assistance Center).

Evaluation of Epidemiological Analysis

As described above, the analyses used spatially aggregated estimates of exposure and of several potential confounders. Thus, the exposure and the confounders vary across at most ~32,000 data points (i.e., zip codes) (and fewer for the covariates aggregated to the county level). That is, the epidemiological analyses presented in the report followed a hybrid study design that mixed characteristics of individual-level cohort studies and of ecological analyses. Though this is not entirely uncommon in this field, a key implication of ecological analyses is that they are unable to capture variability in exposures

or population characteristics present at the individual level. Specifically, one must take the perspective that aggregate exposures (and population characteristics) are equivalent to individual-level exposures (and individual-level characteristics) — for example, that the proportion of low-income individuals in a given zip code represents individual-level poverty. This is not a perfect measure of individual-level poverty, but the investigators argue that at this scale, and as measured, it is adequate for purposes of their analyses.

Another implication for studies based on aggregated data is the potential for the modifiable areal unit problem, in which the observed patterns depend on (and might be biased by) the size and shapes of the arbitrarily defined spatial units of aggregation (i.e., zip codes). Associations between an exposure and health outcome likely operate differently at different scales, and it is not possible to know which scale is most appropriate for any given study. For example, Dominici and colleagues might have found different estimates of risk had they aggregated their data to, say, Census tracts or if the boundaries (shapes or sizes) of the zip codes were defined differently.

The epidemiological analyses are further complicated by the fact that confounders were defined at multiple spatial scales, including some at the individual level (age and sex), others at the zip code level (meteorological variables and indicators of SES), some at the county level (average body mass index and smoking rate), and an indicator for broad regional environment, resulting in a complex hybrid epidemiological model. The Panel felt that this hybrid approach, with confounders measured at several different spatial scales, and in particular with no SES data measured at the individual level, rendered interpretation complicated. On the one hand, when using an aggregated exposure, there cannot be confounding from individual-level variables, although confounding from spatially aggregated values of those variables could be present and was accounted for in the investigators' analyses. On the other hand, aggregation introduces exposure measurement error. The bias from this measurement error is unknown and is difficult to account for statistically, particularly in a complicated real-world analysis, in the context of causal inference, and with multiple pollutants all subject to measurement error.

Notably, in the Phase 1 report, the investigators found that models for $PM_{2.5}$ and cause-specific hospitalization and all-cause mortality were not sensitive to the omission of several individual-level confounders using a nationally representative subsample (~32,000) of Medicare participants with individual information on risk factors. They interpreted those results as an indication that omitting individual risk factors would not lead to biased results in their main analysis. Those findings generally support the validity of the ecological approach to covariate measurement and adjustment presented here. The Panel did note the importance of adjusting for time

in their models as was evident from their sensitivity analyses and were pleased to see year included in the Phase 2 report. They did, however, question whether adjustment for regional environment with only four categories (West, Midwest, South, and Northeast) was sufficient for the purpose of capturing regional variation in unmeasured characteristics that might confound the observed associations.

Regarding the causal analyses, the Panel was impressed by the effort to develop and present three approaches for causal inference that adjusted for confounding using the generalized propensity score by (1) matching, (2) weighting, and (3) adjustment. The Panel was especially pleased with how well the investigators described and defined the assumptions of the generalized propensity score approach and evaluated how well they thought they met the assumptions. That said, the Panel suggests that causal approaches are helpful but are still limited by the underlying data. For example, in this case, the Panel was concerned that applying the causal inference approaches at the zip code level has unclear implications for the statistical properties of the health effects estimation. Ultimately, all approaches are attempting to get at causal relationships, and the key value added in this study was comparing the consistency of findings across multiple approaches.

In summary, the Panel felt that a strength of the report was the collection of epidemiological analyses based on both traditional (i.e., Cox and Poisson) and causal inference approaches. Each approach individually has relative strengths and limitations, but together they allowed the investigators to present a thorough and robust investigation. The Panel felt that interpretation of the results requires a balanced perspective and that it is challenging to assign more weight or value to any one of these results based on the approach alone.

DISCUSSION OF THE FINDINGS AND INTERPRETATION

In this large study with rigorous analyses, including several causal inference approaches, the investigators reported findings that were generally consistent with each other and with those of previous studies. The Panel found it reassuring that the investigators found good consistency in results using five analytical approaches (Commentary Figure 1). It is interesting that models using distinct statistical methods with very different approaches to covariate adjustment all produced effect estimates of generally similar magnitude (i.e., HRs for $PM_{2.5}$ on all-cause mortality all between 1.06 and 1.08 per $10 \mu\text{g}/\text{m}^3$). However, such a result is not wholly unexpected, given that the analyses were all conducted with the exact same datasets.

The Panel appreciated that Dominici and colleagues presented results from all five statistical methods for the full cohort and the low-exposure cohort. The latter analyses in particular contributed important evidence of effects on health associated with relatively low concentrations of ambient pollution. The findings contributed to the small, yet increasing, body of evidence reporting adverse health effects associated

with exposures to such low concentrations of ambient pollution.

The Medicare cohort used in the study consists of older Americans (ages 65 and over at baseline, mean age 69.2 years). The Panel was uncertain about the generalizability of the findings presented here to other age groups or to those living in other geographic locations. For example, it is not clear to what extent the risks estimated here for older adults might compare with those for younger adults. This issue was not discussed in the report.

The presentation of ER functions for both single and multipollutant models was another important contribution of the report. The presentation format of the figures was clear, and it was helpful to be able to compare the single- and multipollutant figures next to each other. As noted above (and shown in Commentary Figure 2), the plots showed evidence of associations between mortality and long-term exposures to $PM_{2.5}$ as low as $3 \mu\text{g}/\text{m}^3$. In the case of $PM_{2.5}$, the shapes of the ER functions were almost linear. It is important to note here that the investigators emphasized that they drew their main conclusions for the study from the single-pollutant models and that it remains unclear whether ambient NO_2 or O_3 actually serve as confounders of the relationships between ambient $PM_{2.5}$ and health outcomes. Although the Panel would agree on this point, it nevertheless leaves open the possibility of confounding by copollutants in single-pollutant analyses.

Regarding the overall interpretation of the causal inference models for $PM_{2.5}$ and mortality, the Panel appreciated that the investigators did not overextend their confidence in the results of the models in demonstrating causality.

CONCLUSIONS

In summary, this study represents an important contribution to the literature on the health effects of long-term exposure to ambient air pollution in a very large cohort of older adults in the United States. Dominici and colleagues conducted an extensive and innovative set of analyses, including traditional regression models and causal inference models, with very large air pollution and health data sets. They reported evidence from their causal inference analyses of relationships between mortality and long-term exposures to $PM_{2.5}$ and NO_2 . For O_3 , ER functions with all-cause mortality were almost flat below 45 ppb and showed no statistically significant effects, but there was evidence of increased hazard at exposures greater than 45 ppb. Moreover, the estimates of mortality risk associated with $PM_{2.5}$ exposure were generally similar using the five different statistical approaches and remained elevated among participants with longer-term exposures below or equal to $12 \mu\text{g}/\text{m}^3$, the current NAAQS for $PM_{2.5}$.

The effect estimates reported here for $PM_{2.5}$ on all-cause mortality were similar to those reported in several previous studies that have considered these associations at low

exposures. In their work, Dominici and colleagues have used a massive dataset of mortality from older adults across the full United States over more than 15 years. With their spatial prediction models and causal modeling approaches, they have overcome some of the limitations of previous studies. However, the complex hybrid nature of the analyses — in which they used several spatial scales across the many variables included — makes it difficult to understand fully the implications of these hybrid approaches. Thus, there remain some potential sources of error that could have affected the results. These include (1) the likely greater error in estimating rural concentrations due to the relative paucity of ground monitors for evaluation and training of exposure models in those areas, (2) the exposure measurement error from using zipcode aggregated exposure estimates, and (3) the effects of using aggregated covariates (at several spatial scales) in adjusting for confounding. Ultimately, the major contribution of this study is that using several different approaches, the investigators produced findings that were generally consistent with each other and with those of previous studies.

ACKNOWLEDGMENTS

The HEI Review Committee is grateful to the Low-Exposure Epidemiology Studies Review Panel for its thorough review of the study. The Committee is also grateful to Hanna Boogaard for her oversight of the study, to Eva Tanner for her assistance in reviewing the report, to Dan Crouse and Martha Ondras for their assistance in reviewing the report and in preparing its Commentary, to George Simonson for his editing of the report and its Commentary, and to Hope Green and Kristin Eckles for their roles in preparing the report for publication.

REFERENCES

- Beelen R, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, Hoffmann B, et al. 2014a. 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; doi:10.1016/S0140-6736(13)62158-3.
- Beelen R, Stafoggia M, Raaschou-Nielsen O, Andersen ZJ, Xun WW, Katsouyanni K, et al. 2014b. Long-term exposure to air pollution and cardiovascular mortality: An analysis of 22 European cohorts. *Epidemiology* 25:368–378; doi:10.1097/EDE.0000000000000076.
- Chen J, Hoek G. 2020. Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. *Environ Int* 143:105974; doi:10.1016/j.envint.2020.105974.
- Chen T, Guestrin C. 2016. Xgboost: A scalable tree boosting system. In *KDD '16: Proceedings of the 22nd ACM SIGKDD*

- International Conference on Knowledge Discovery and Data Mining. Available: <https://doi.org/10.1145/2939672.2939785> [accessed 25 August 2021].
- Coughlin SS, Clary C, Johnson JA, Berman A, Heboyan V, Ben-vides T, et al. 2019. Continuing challenges in rural health in the United States. *J Environ Health Sci* 5:90–92.
- Crouse DL, Peters PA, Hystad P, Brook JR, van Donkelaar A, Martin RV, et al. 2015. Ambient PM_{2.5}, O₃, and NO₂ exposures and associations with mortality over 16 years of follow-up in the Canadian census health and environment cohort (Can-CHEC). *Environ Health Perspect* 123:1180–1186; doi:10.1289/ehp.1409276.
- Crouse DL, Peters PA, van Donkelaar A, Goldberg MS, Ville-neuve 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; doi:10.1289/ehp.1104049.
- Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, et al. 2019. An ensemble-based model of PM_{2.5} concentration across the contiguous United States with high spatiotemporal resolution. *Environ Int* 130:104909; doi:10.1016/j.envint.2019.104909.
- Di Q, Amini H, Shi L, Kloog I, Silvern R, Kelly J, et al. 2020. Assessing NO₂ concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. *Environ Sci Technol* 54:1372–1384; doi:10.1021/acs.est.9b03358.
- 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; doi:10.1001/jama.2017.17923.
- Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. 2017b. Air pollution and mortality in the Medicare population. *N Engl J Med* 376:2513–2522; doi:10.1056/NEJ-Moa1702747.
- 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.
- Global Burden of Disease (GBD) 2019 Risk Factors Collaborators. 2020. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 396:1223–1249; doi:10.1016/S0140-6736(20)30752-2.
- Hales S, Blakely T, Woodward A. 2012. Air pollution and mortality in New Zealand: Cohort study. *J Epidemiol Community Health* 66:468–473; doi:10.1136/jech.2010.112490.
- Health Effects Institute. 2020. State of Global Air 2020. Special Report. Boston, MA:Health Effects Institute.
- Hirano K, Imbens GW. 2004. The propensity score with continuous treatments. In: Applied Bayesian Modeling and Causal Inference from Incomplete Data Perspectives (Gelman A, Meng X-L, eds). Hoboken, NJ:John Wiley & Sons, Ltd.
- Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. 2011. Assessing temporally and spatially resolved PM_{2.5} exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos Environ* 45:6267–6275; doi:10.1016/j.atmosenv.2011.08.066.
- O’Neill MS, Jerrett M, Kawachi I, Levy JI, Cohen AJ, Gouveia N, et al. 2003. Health, wealth, and air pollution: Advancing theory and methods. *Environ Health Perspect* 111:1861–1870; doi:10.1289/ehp.6334.
- Papadogeorgou G, Dominici F. 2020. A causal exposure response function with local adjustment for confounding: Estimating health effects of exposure to low levels of ambient fine particulate matter. *Ann Appl Stat* 14:850–871; doi:10.1214/20-aos1330.
- Pappin AJ, Christidis T, Pinault LL, Crouse DL, Tjepkema M, Erickson AC, et al. 2019. Examining the shape of the association between low levels of fine particulate matter and mortality across three cycles of the Canadian Census Health and Environment Cohort. *Environ Health Perspect* 127:10; doi:10.1289/EHP5204.
- Pinault L, Tjepkema M, Crouse DL, Weichenthal S, van Donkelaar A, Martin RV, et al. 2016. Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. *Environ Health* 15:18–31; doi:10.1186/s12940-016-0111-6.
- Requia WJ, Di Q, Silvern R, Kelly J, Koutrakis P, Mickley LJ, et al. 2020. An ensemble learning approach for estimating high spatiotemporal resolution of ground level ozone in the contiguous United States. *Environ Sci Technol* 54:11037–11047; doi:10.1021/acs.est.0c01791.
- Robins JM, Hernan MA, Brumback B. 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11:550–560; doi:10.1097/00001648-200009000-00011.
- U.S. EPA (Environmental Protection Agency). 2004. Vol I. PM Air Quality Criteria Document. Washington, DC:U.S. Environmental Protection Agency.
- U.S. EPA (Environmental Protection Agency). 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 [accessed 24 April 2019].
- U.S. EPA (Environmental Protection Agency). 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. *Ann Appl Stat* 13:520–547; doi:10.1214/18-AOAS1206.

Wu X, Mealli F, Kioumourtzoglou MA, Dominici F, Braun D. In review. Matching on generalized propensity scores with continuous exposures. Available: <https://arxiv.org/abs/1812.06575> [accessed 25 August 2021].

Zhu Y, Coffman DL, Ghosh D. 2015. A boosting algorithm for estimating generalized propensity scores with continuous treatments. *J Causal Inference* 3:25–40; doi:10.1515/jci-2014-0022.

RELATED HEI PUBLICATIONS

Number	Title	Principal Investigator	Date
Research Reports			
208	Mortality and Morbidity Effects of Long-Term Exposure to Low-Level PM _{2.5} , BC, NO ₂ , and O ₃ : An Analysis of European Cohorts in the ELAPSE Project	B. Brunekreef	2021
203	Mortality–Air Pollution Associations in Low-Exposure Environments (MAPLE): Phase 1	M. Brauer	2019
200	Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1	F. Dominici	2019
193	Particulate Air Pollutants, Brain Structure, and Neurocognitive Disorders in Older Women	J-C. Chen	2017
192	Multicenter Ozone Study in older Subjects (MOSES): Part 1. Effects of Exposure to Low Concentrations of Ozone on Respiratory and Cardiovascular Outcomes	M.W. Frampton	2017
187	Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations	C.M. Zigler	2016
183	Development of Statistical Methods for Multipollutant Research	B.A. Coull	2015
142	Air Pollution and Health: A European and North American Approach	K. Katsouyanni	2009
140	Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality	D. Krewski	2009
139	Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: The NLCS-AIR Study	B. Brunekreef	2009
Special Report			
	Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality	Health Effects Institute	2000

ABBREVIATIONS AND OTHER ITEMS

AC	absolute correlation
BMI	body mass index
BRFSS	Behavioral Risk Factor Surveillance System
CanCHEC	Canadian Census Health and Environment Cohorts
CI	confidence interval
ELAPSE	Effects of Low-Level Air Pollution: A Study in Europe
ER	exposure response
ERC	exposure–response curve
GPS	generalized propensity score
HR	hazard ratio
IR	investigators' report
LERCA	local exposure response confounding adjustment
MAPLE	Mortality–Air Pollution Associations in Low-Exposure Environments
MBSF	master beneficiary summary file
MCBS	Medicare Current Beneficiary Survey
NAAQS	National Ambient Air Quality Standards
NO ₂	nitrogen dioxide
O ₃	ozone
PI	principal investigator
PM	particulate matter
PM _{2.5}	particulate matter ≤2.5 μm in aerodynamic diameter
ppb	parts per billion
RC	regression calibration
RFA	request for applications
SES	socioeconomic status
TEA	total events avoided
U.S. EPA	U.S. Environmental Protection Agency

HEI BOARD, COMMITTEES, and STAFF

Board of Directors

Richard A. Meserve, Chair *Senior of Counsel, Covington & Burling LLP; President Emeritus, Carnegie Institution for Science; former Chair, U.S. Nuclear Regulatory Commission*

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*

Karen C. Seto *Frederick Hixon Professor of Geography and Urbanization Science, Yale School of the Environment*

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*

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

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

Heather A. Holmes *Associate Professor, Department of Chemical Engineering, University of Utah*

Neil Pearce *Professor of Epidemiology and Biostatistics, London School of Hygiene and Tropical Medicine*

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

Evangelia (Evi) Samoli *Associate Professor of Epidemiology and Medical Statistics, Department of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, Greece*

Gregory Wellenius *Professor, Department of Environmental Health, Boston University School of Public Health*

Review Committee

Melissa Perry, Chair *Professor and Chair, Department of Environmental and Occupational Health, George Washington University Milken Institute School of Public Health*

Sara D. Adar *Associate Professor and Associate Chair, Department of Epidemiology, University of Michigan School of Public Health*

Kiros Berhane *Professor and Chair, Department of Biostatistics, Mailman School of Public Health, Columbia University*

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

Frank Kelly *Henry Battcock Chair of Environment and Health and Director of the Environmental Research Group, Imperial College London School of Public Health*

HEI BOARD, COMMITTEES, and STAFF

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*

Eric J. Tchetgen Tchetgen *Luddy Family President's Distinguished Professor, Professor of Statistics and Data Science, The Wharton School, University of Pennsylvania*

Officers and Staff

Daniel S. Greenbaum *President*

Robert M. O'Keefe *Vice President*

Ellen K. Mantus *Director of Science*

Donna J. Vorhees *Director of Energy Research*

Annemoon M. van Erp *Deputy Director of Science*

Thomas J. Champoux *Director of Science Communications*

Jacqueline C. Rutledge *Director of Finance and Administration*

Emily Alden *Corporate Secretary*

Lee Ann Adelsheim *Research Assistant*

Palak Balyan *Consulting Staff Scientist*

Hanna Boogaard *Consulting Principal Scientist*

Sofia Chang-DePuy *Digital Communications Manager*

Aaron J. Cohen *Consulting Principal Scientist*

Dan Crouse *Senior Scientist*

Robert M. Davidson *Staff Accountant*

Philip J. DeMarco *Compliance Manager*

Kristin C. Eckles *Senior Editorial Manager*

Elise G. Elliott *Staff Scientist*

Hope Green *Editorial Project Manager*

Joanna Keel *Research Assistant*

Yi Lu *Staff Scientist*

Lissa McBurney *Senior Science Administrator*

Janet I. McGovern *Executive Assistant*

Martha Ondras *Research Fellow*

Pallavi Pant *Senior Scientist*

Allison P. Patton *Senior Scientist*

Quoc Pham *Science Administrative Assistant*

Anna S. Rosofsky *Staff Scientist*

Robert A. Shavers *Operations Manager*

Eva Tanner *Staff Scientist*

Eleanne van Vliet *Staff Scientist*



HEALTH EFFECTS INSTITUTE

75 Federal Street, Suite 1400

Boston, MA 02110, USA

+1-617-488-2300

www.healtheffects.org

RESEARCH REPORT

Number 211

January 2022

