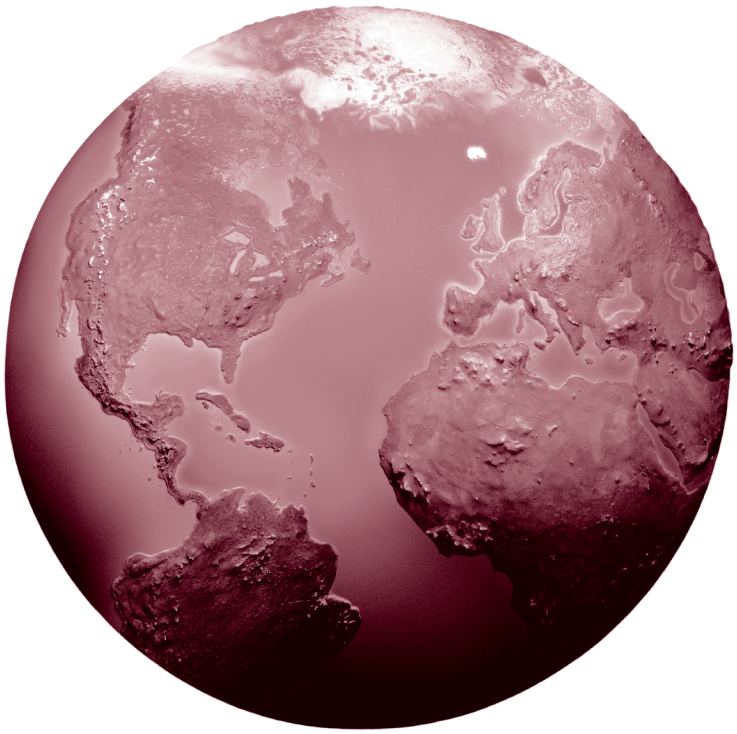
HEI

January 2022



COMMENTARY BY THE HEI LOW-EXPOSURE EPIDEMIOLOGY STUDIES REVIEW PANEL SUMMARIZING AND EVALUATING THE INVESTIGATORS' REPORT:

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

Dominici et al.

Health Effects Institute

Commentary by the HEI Low-Exposure Epidemiology Studies Review Panel Summarizing and Evaluating the Investigators' Report:

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

Research Report 211

Health Effects Institute

Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

CONTENTS

ABOUT HEI

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

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

HEI typically receives balanced funding from the U.S. Environmental Protection Agency and the worldwide motor vehicle industry. Frequently, other public and private organizations in the United States and around the world also support major projects or research programs. HEI has funded more than 340 research projects in North America, Europe, Asia, and Latin America, the results of which have informed decisions regarding carbon monoxide, air toxics, nitrogen oxides, diesel exhaust, ozone, particulate matter, and other pollutants. These results have appeared in more than 260 comprehensive reports published by HEI, as well as in more than 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.

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*

COMMENTARY
Low-Exposure Epidemiology Studies Review Panel

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;

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.

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

- Conducting risk and policy assessments to estimate public 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 [NO_2], and ozone [O_3]) 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 $PM_{2.5}$, NO_2, and O_3 are as follows (*https://www.epa.gov/criteria-air-pollutants/naaqs-table*):

- $PM_{2.5}$: annual mean averaged over 3 years of 12 μg/m^3;
- NO_2: annual mean of 53 ppb (approximately 100 μg/m^3); and
- O_3: annual fourth-highest daily maximum 8-hour concentration, averaged over 3 years, of 70 ppb (approximately 140 μg/m^3).

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/m^3. 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 NO_2 or O_3. 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 12 $\mu g/m^3$ 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 × 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	• Summarized exposure predictions for daily $PM_{2.5}$ and O_3 at 1-km × 1-km grid for the contiguous United States over the period 2000–2012, using an ensemble modeling approach (Di et al. 2019).	• 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	• 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.	• 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	• 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.	• 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	• Provided code for implementation of new causal inference methods.	• 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 km^2 in Los Angeles County, California, and 268 km^2 in the state of Texas. In total, there are about 42,000 zip codes in the United States, with a mean area of 234 km^2, 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-km^2 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-km^2 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 $\mu g/m^3$ during every year of follow-up (i.e., the low-exposure cohort described earlier). This exposure cut point was selected because 12 $\mu g/m^3$ 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 μg/m³ (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-μg/m³ 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 μg/m³ than among all studies and among those conducted at mean concentrations below 25 μg/m³ (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 12 μg/m³, 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.

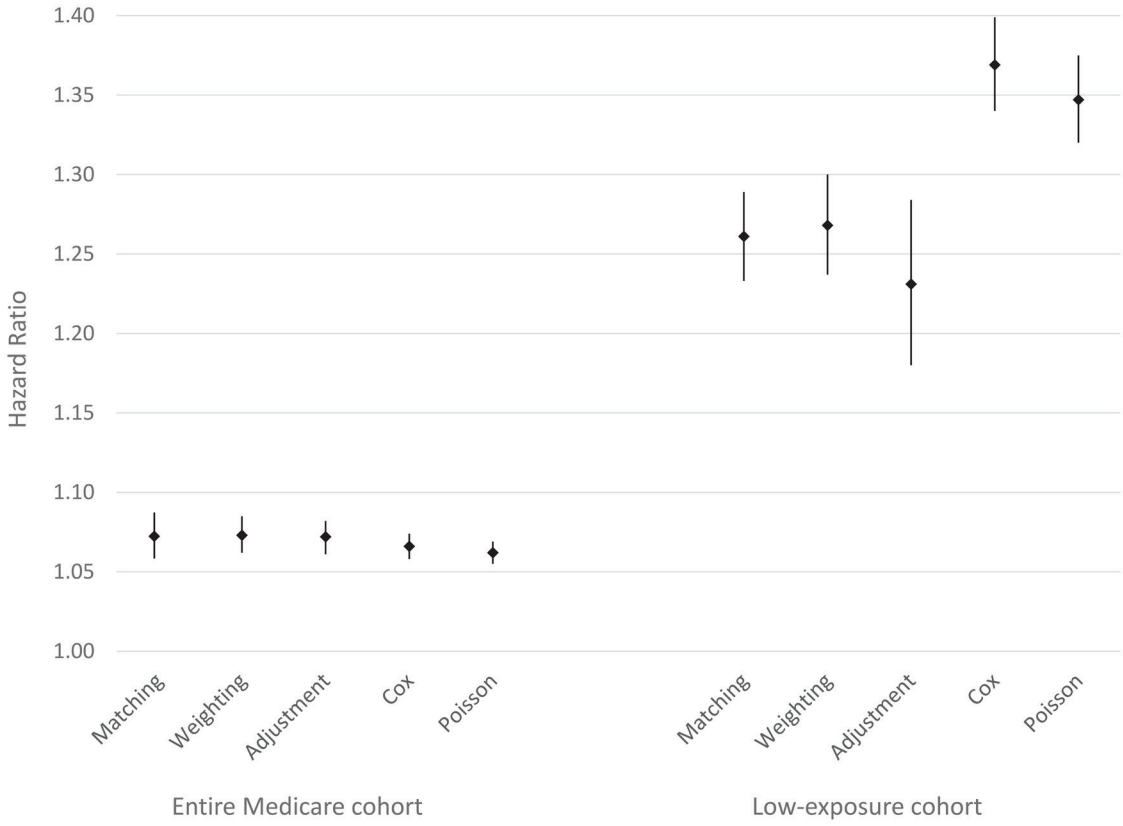
Exposure–Response Functions

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 μg/m³, 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 nonlinear 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-μg/m³ increase in $PM_{2.5}$ exposure. Results are presented for fully adjusted models. (Source: Adapted from Figure 6 in the Investigators' Report.)

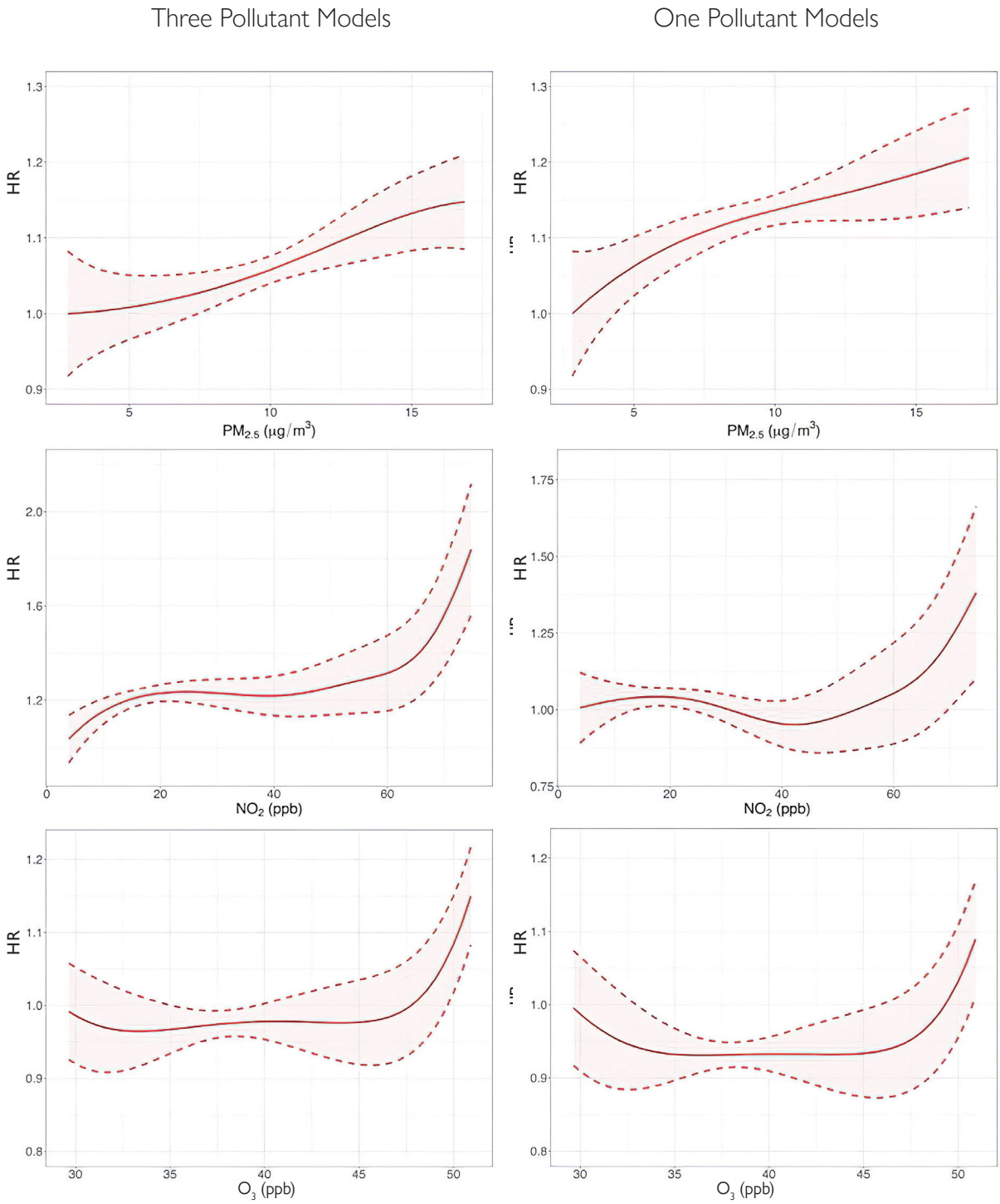
Reproducible Research

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 μg/m³. They reported consistent increases in risk of all-cause mortality ranging from 6% to 8% per 10 μg/m³ 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 km^2 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 $\mu g/m^3$). 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 air 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 air 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 μg/m³. 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 12 μg/m³, 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, Benevides 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 $PM_{2.5}$, O_3, and NO_2 exposures and associations with mortality over 16 years of follow-up in the Canadian census health and environment cohort (CanCHEC). Environ Health Perspect 123:1180–1186; doi:10.1289/ehp.1409276.

Crouse DL, Peters PA, van Donkelaar A, Goldberg MS, Villeneuve PJ, Brion O, et al. 2012. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. Environ Health Perspect 120:708–714; 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 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; 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/NEJMoa1702747.

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-aoas1330.

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. In review.

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.

ABBREVIATIONS AND OTHER ITEMS

CI	confidence interval
ER	exposure–response
HR	hazard ratio
IR	investigators' report
NAAQS	National Ambient Air Quality Standards
NO_2	nitrogen dioxide
O_3	ozone
PM	particulate matter
$PM_{2.5}$	particulate matter ≤2.5 μm in aerodynamic diameter
RFA	request for applications
SES	socioeconomic status
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

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.