Health Effects Institute
75 Federal Street
Suite 1400
Boston MA 02110 USA
+1-617-488-2300
FAX +1-617-488-2335
www.healtheffects.org

COMMENTS OF THE HEALTH EFFECTS INSTITUTE
ON THE DRAFT INTEGRATED SCIENCE ASSESSMENT FOR OZONE
DOCKET ID: EPA-HQ-ORD-2018-0274

December 2, 2019

The Health Effects Institute (HEI) is pleased to have the opportunity to submit these comments to US EPA and the Clean Air Scientific Advisory Committee (CASAC) on the new draft Integrated Science Assessment for ozone. We are encouraged to see that the draft ozone ISA continues to adhere to the high standards of scientific quality and systematic review of the literature which has become a hallmark of EPA's approach to meeting the requirements of Section 108 of the Clean Air Act. That section of the Act requires that such reviews "shall accurately reflect the latest scientific knowledge useful in indicating the kind and extent of all identifiable effects on public health or welfare which may be expected from the presence of such (*criteria*) pollutant in the ambient air."

As you are aware, HEI has produced a number of studies of air pollution and health to inform the review of the National Ambient Air Quality Standards (NAAQS) and we were pleased to see several of these studies cited in the draft ISA. We will not review the specific studies (though we stand ready to answer any questions that CASAC might have on them). Rather we wanted to focus in these comments on:

1. The causality determinations in the draft ozone ISA, and
2. EPA conclusions of several key health endpoints where HEI studies have contributed importantly to the evidence

1. Causality Determinations HEI has followed closely the development and application of EPA's criteria for assessing causality of different air pollutants on particular health outcomes since they were first applied in the review of the NAAQS for NOx in 2008. We have found that this approach has been enhanced significantly over previous reviews, especially because it includes:

- Well stated criteria for causality determination presented *a priori* in the Preface of each ISA;
- Careful evaluation of evidence from multiple disciplines: exposure assessment, toxicology, clinical studies, and epidemiology, rather than reliance on any one strand of evidence or solely on statistical causal analyses; and
- Explicit acknowledgement of the uncertainties attendant in each case.

The result of this process is an open presentation of the literature and assumptions applied, and the opportunity for both CASAC and the broader community to review and raise questions about the determinations.

Given this opportunity, HEI has reviewed the latest determinations in the draft ozone ISA, which in general seem well and carefully done. We do have comments on the certain determinations summarized in the ISA, broadly agreeing with their conclusions but offering some remarks.

Evidence for short term respiratory effects from chamber studies: The ISA summarizes the literature on short term respiratory effects, specifically on changes in measures of pulmonary function, which has bene widely reported and which has played a central role in establishment of the ozone standard. The unresolved question has been at how low a level can such effects be observed and in what populations. The HEI-supported MOSES study has provided useful information and we are pleased to see it cited in the ISA, and that it helps strengthen the determination of a causal relationship for the respiratory effects (See HEI MOSES report and investigator publications).

In HEI's Research Report 192, *Multicenter Ozone Study in oldEr Subjects (MOSES): Part 1. Effects of Exposure to Low Concentrations of Ozone on Respiratory and Cardiovascular Outcomes*, 87 healthy subjects, average age 60, were exposed to three levels of ozone, namely clean air control, and 70 and 120 ppb of ozone. This multicenter study incorporated extensive quality assurance and control procedures and measured a wide variety of end points. Ozone exposure caused concentration-related reductions in lung function and presented evidence for airway inflammation and injury. Though a couple of other studies have made observed similar effects at low levels, the MOSES study stands out as having exposed the largest number of subjects, under well characterized conditions, and to some of the lowest ozone concentrations. And these observations are all the more noteworthy because healthy older individuals are less responsive to ozone-induced lung function effects than are healthy young individuals

Lack of Evidence for Cardiovascular Effects after Short-term Exposure. The question of the potential cardiovascular effects of ozone has been an important concern. We agree with the overall conclusion in the ISA that lowers the strength of evidence to "suggestive of a causal relationship" from the previous "Likely to be a causal relationship." Again, we are pleased to see that the HEI MOSES study has provided useful information for this endpoint as well.

HEI's primary motivation for HEI's MOSES study was to evaluate whether short-term exposure of older, healthy individuals to ambient levels of ozone induces acute cardiovascular responses." Using a cross-over design, this multicenter study efficiently collected information on a comprehensive array of cardiovascular endpoints, probing a variety of potential mechanistic or pathophysiological pathways, as well as several respiratory endpoints, and found that a 3-hour ozone exposure at 70 or 120 ppb did not lead to statistically significant changes in cardiovascular endpoints in this healthy group of 87 older participants undergoing moderate exercise.

Exposure to Ozone and Mortality: The draft ozone ISA reviews a wide range of newer epidemiologic studies of ozone exposure and health, an area where study results have not been consistent, and concludes that the overall evidence is "suggestive of, but not sufficient to infer, a causal relationship." One of the considerations in this context is whether there is evidence for robust associations at levels of exposure below the current NAAQS, i.e. 70 ppb. While there had been a few earlier studies reporting such associations, these were relatively small in number, and questions remained about potential exposure measurement error at the lowest levels and the absence of information on potential confounders. These issues are being addressed in three HEI funded studies under our Health Effects at Low Levels of Air Pollution program, particularly a study based on the Medicare population in the US, whose results have been published in the open literature and cited in the ISA.

After a rigorous peer-review of the results of the first two years of this research project, HEI has very recently published these early results, along with a commentary, prepared by a specially appointed review panel (Research Report 200, *Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1*, Dominici et al). Briefly, Francesca Dominici and her colleagues have analyzed data from 61 million Medicare enrollees, between the years 2000 and 2012, using exposure estimates based on sophisticated hybrid models with a resolution of 1 km^2 which included populations living in less well monitored areas. The initial results of these analyses – for both long and short term exposures – suggest an association between with all-cause mortality in two pollutant analysis – including at levels below the current ozone NAAQS. Substantial additional work is underway examine exposure measurement error, explore the role of various confounders, and to test causal inference models. The analyses are also being extended to 2016, making this study using not only the largest population but also one using the most recent air quality and health data.

We believe that the evidence for an association between ozone exposure and an increase in mortality is getting stronger. However, the additional work HEI is supporting under its Low Exposure program and other studies would help shed light on the remaining uncertainties as well as clarify the nature of such relationship.

HEI is attaching the Statements and Commentaries for this report to these Comments on the Ozone ISA and will respectfully request that CASAC, recognizing that this involves further intensive HEI review of studies already cited in the draft ozone ISA, would consider including this HEI Research Reports 200 in the final version of the ozone ISA.

CONCLUSION

Thank you again for the opportunity to provide these comments. We would be pleased to provide any additional assistance to EPA and CASAC in its review process, and/or answer any additional questions you may have about these comments, or the range of other HEI Research Reports under consideration in the Draft ozone ISA.

HEI

ATTACHMENT 1

Statement and Commentary on HEI Research Report 200

by HEI's Low-Exposure Epidemiology Studies Review Panel

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

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

†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.

Health Effects Institute, Boston, Massachusetts

CONTENTS

Statement

by HEI's Low-Exposure Epidemiology Studies Review Panel

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

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

What This Study Adds

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

INTRODUCTION

The levels of most ambient air pollutants have declined significantly in the United States during the last few decades. Recent epidemiological studies, however, have suggested an association between exposure to ambient levels of air pollution — even below the current U.S. National Ambient Air Quality Standards (NAAQS) — and adverse health effects. In view of the importance of such research findings, the Health Effects Institute in 2014 issued a request for applications (RFA 14-3), seeking to fund research to assess the health effects of long-term exposure to low levels, particularly below the NAAQS, of ambient air pollution and to develop improved statistical methods for conducting such research. HEI funded three studies under this program; each study used state-of-the-art exposure methods and very large cohorts. The studies were based in the United States, Canada, and Europe, thus providing a comprehensive cross-section of high-income countries where ambient levels are generally low.

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

APPROACH

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

Data and Methods: Dominici and colleagues developed hybrid air pollution concentration models for the contiguous United States for the period 2000 to 2012, using data from a variety of sources,

including satellite data, chemical transport models, land-use and weather variables, and routinely collected air monitoring data from the U.S. EPA.

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

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

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

Overall	
Medicare – Study Population	60.9 million
MCBS Population	57,200
Study Period	2000–2012
Case–Control Study	
Follow-up period	460.3 million person-years
Deaths	22.6 million
$PM_{2.5}$ Average Concentration	11.0 μg/m^3
Ozone Average Concentration	46.3 ppb
Case–Crossover Study	
Case Days	22.4 million
Control Days	76.1 million
$PM_{2.5}$ Average Concentration	11.6 μg/m^3
O_3 Average Concentration	37.8 ppb

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

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

RESULTS

Dominici and colleagues report overall good performance of the models for estimating $PM_{2.5}$ and O_3 concentrations, with overall R^2 values of 0.84 and 0.80, respectively. For $PM_{2.5}$, the average annual concentration was 11.0 $\mu g/m^3$ during the study period, 2000–2012. Performance of the model varied between different geographical regions and seasons; the highest $PM_{2.5}$ concentrations were predicted to be in California and the eastern and southeastern United States, and model performance was better in the eastern and central United States than in the western part of the country. And, the $PM_{2.5}$ model performed best during the summer. For O_3, the average of 8-hour-daily concentrations

during the warm season was 46.3 ppb during the study period. O_3 concentrations were highest in the Mountain region and in California and lower in the eastern states. The average concentrations of $PM_{2.5}$ decreased during the study period, but O_3 concentrations remained more or less the same. Annual $PM_{2.5}$ and warm-season O_3 concentrations were only weakly correlated.

The 2000–2012 cohort of Medicare beneficiaries provided a very large population for studying association with long-term effects of exposure to ambient air pollution. In two-pollutant analyses of long-term effects, Dominici and colleagues report a 7.3% higher risk of all-cause mortality for each 10-μg/m^3 increase in annual average $PM_{2.5}$ concentrations and a 1.1% higher risk of mortality for each 10-ppb increase in average O_3 concentrations in the warm season. At low concentrations — less than 12 μg/m^3 $PM_{2.5}$ and less than 50 ppb O_3 — the risk was 13.6% for $PM_{2.5}$ and 1.0% for O_3 for each 10 μg/m^3 and 10 ppb increase in concentrations, respectively The concentration–response relationships from the two-pollutant models showed almost linear curves, with no suggestion of a threshold down to 5 μg/m^3 $PM_{2.5}$ and 30 ppb O_3.

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

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

INTERPRETATION OF RESULTS

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

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

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

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

The greatest challenge to the internal validity of this study, as for all observational studies, is the potential for confounding, which can bias the results. To address such concerns, the investigators

performed numerous analyses with some 20 covariates. They also utilized findings from a smaller Medicare cohort that had a much richer set of potential confounding variables to assess the likely impact of having only a limited number of covariates in the main cohort analysis. In addition, to allow for the effects of time-dependent covariates known to vary from year to year, they utilized a variant of the classic Cox proportional hazards model, the Andersen-Gill formulation.

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

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

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

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

pollutants — such as NO_2 — may also confound the associations between $PM_{2.5}$ and O_3 and mortality.

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

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

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

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

In the first method, the investigators have developed a generalized-propensity-score approach for confounding adjustment along with a regression calibration method to address exposure measurement error in health models. In the second approach, they have developed a new Bayesian

causal approach, known as local exposure–response confounding adjustment, to estimate exposure–response curves accounting for differential effect of confounders at different levels of exposure. Both of these approaches serve as potentially useful starting points, and the Panel notes that current applications do not address the concerns raised about the long-term and short-term studies — in particular, concerns about residual confounding and impacts of the complex hybrid nature of the study designs — and so it looks forward to the full development and applications of these methods to the health analyses.

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

CONCLUSIONS

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

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

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

Commentary

by HEI's Low-Exposure Epidemiology Studies Review Panel

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

INTRODUCTION

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

SCIENTIFIC AND REGULATORY BACKGROUND

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

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

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

The U.S. CAA requires that in setting the National Ambient Air Quality Standards (NAAQS), the U.S. Environmental Protection Agency (U.S. EPA) Administrator review all available science and set the NAAQS for all major (the "criteria") pollutants (including ozone [O_3], particulate matter

[PM], and nitrogen dioxide [NO_2]) at a level "requisite to protect the public health with an adequate margin of safety." In practice, since 2008 that review has had two principal steps:

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

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

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

EVOLUTION OF THE NAAQS

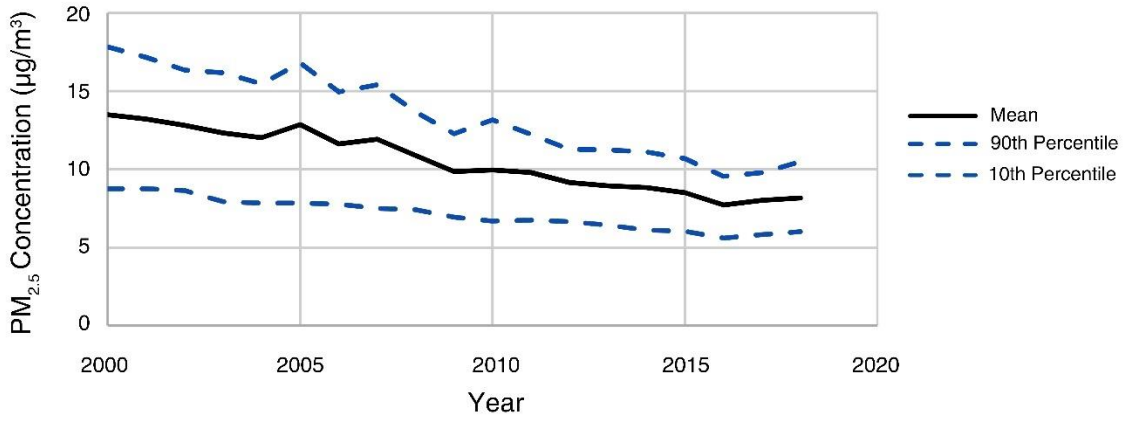
The reviews of the criteria pollutants have been ongoing for nearly 50 years, since the passage of the Clean Air Act Amendments of 1970. As the science has evolved, each subsequent review has examined the strength of the evidence for retaining or tightening the NAAQS. Although the process has frequently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

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

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

IMPACT OF THE NAAQS

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



Commentary Figure. Trends in $PM_{2.5}$ concentration from 2000 to 2018 (seasonally weighted annual average) as monitored by the U.S. EPA (data from U.S. EPA).

ADVENT OF RECENT STUDIES OBSERVING ASSOCIATIONS BELOW THE NAAQS

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

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

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

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

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

SUMMARY OF THE STUDY

SPECIFIC AIMS

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

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

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

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

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

This commentary focuses in more detail on Aims 1 and 3, comprising initial results from the exposure and health effects research that have been published in peer-reviewed journals (Di et al. 2017a, 2017c). Aims 2 and 4 (causal modeling and more advanced statistical analyses and data access) are discussed briefly here as this research is still in its initial stages; the Panel does offer some comments on this research, with suggestions for the conduct of these further analyses.

EXPOSURE AND HEALTH EFFECTS STUDIES

Data and Methods

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

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

1. Air monitoring data were obtained from the U.S. EPA Air Quality System (AQS), used in both model building and for cross-validation.
2. Aerosol optical depth (AOD) data were obtained from Moderate Resolution Imaging Spectroradiometer (MODIS).
3. Surface reflectance data were also obtained from MODIS (MOD09A1).
4. Chemical transport model outputs were derived from the widely used GEOS-Chem model, which uses meteorological inputs and emission inventories to simulate atmospheric components. Total $PM_{2.5}$ was defined as the sum of nitrate, sulfate and ammonium ions and elemental carbon, organic carbon, sea salt aerosol, and dust aerosol. In addition to producing ground-level $PM_{2.5}$ estimates, the GEOS-Chem model is also useful for calibrating AOD because, being a three-dimensional model, it simulates vertical distribution of aerosols.
5. Meteorological data were obtained from the North American Regional Reanalysis project; the variables used included air temperature, accumulated total precipitation, downward shortwave radiation flux, accumulated total evaporation, planetary boundary layer height, low cloud area

fraction, precipitable water for the entire atmosphere, pressure, specific humidity at 2 meters, visibility, wind speed, medium cloud area fraction, high cloud area fraction, and surface reflectance.

6. Aerosol index data were taken from the absorbing aerosol index measured by the ozone monitoring instrument (OMI), onboard the Aura satellite. These data are used to correct for the presence of other absorbing aerosols in the air (such as those from biomass burning and desert dust).
7. Land-use terms were obtained as previously described by Kloog and colleagues (2012). These terms represent emissions and can help inform small spatial scale variations; land-use data incorporate a variety of variables (such as population and road densities, emissions inventory, elevation, percentage urban, etc.).
8. In the regression models, the investigators also used regional and dummy variables to account for regional and temporal variability due to differences in meteorology and aerosol composition.

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

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

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

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

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

Commentary Table 1. Characteristics of potential confounders and variables*

Potential Confounder	Model Covariate	Variable Level	Variable type	Data Source**
Age	Age at entry	Individual	Categorical (5 year)	Medicare
Race	White Black (%) Asian Hispanic (%) Native American	Individual	Binary and Continuous (% of population)	Medicare and US Census, ACS
Sex	Sex	Individual	Binary	Medicare
Smoking	Ever Smoker (%)	Ecologic (county to Zip code)	Proportion	BRFSS (2000-2012)
Obesity	BMI	Ecologic (county to Zip code)	Continuous	BRFSS (2000-2012)
Diet	Not included	n/a	n/a	n/a
Exercise	Not Included	n/a	n/a	n/a
Socioeconomic status -- Individual level	Medicaid Eligibility	Individual	Binary	Medicaid Statistical Information System
Socioeconomic Status – Community level	Median Household Income	Ecologic (Zip code)	Continuous	US Census, ACS
	Median value of housing	Ecologic (Zip code)	Continuous	US Census, ACS
	% owner occupied	Ecologic (Zip code)	Continuous	US Census, ACS
	% below poverty level (age>65)	Ecologic (Zip code)	Continuous	US Census, ACS
	% below high school education (age>65)	Ecologic (Zip code)	Continuous	US Census, ACS
	Population Density	Ecologic (Zip code)	Continuous	US Census, ACS
Access to Health Care	% with LDL-C % with HgbA1c test % with $\geq$ 1 visit	Ecologic (Zip code)	Continuous	Dartmouth Atlas of Health Care
Meteorological	Temperature Relative humidity	Area (32 km × 32 km)	Continuous	North American Regional Reanalysis data
Regional Dummy Variable	10 geographical regions with similar PM2.5 chemical profile	Region	Categorical	GEOS-Chem 3D global chemical transport model

*Based on information in Di et al 2017c, Supplementary Materials.

**American Community Survey (ACS); Behavioral Risk Factor Surveillance System (BRFSS)

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

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

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

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

The investigators controlled for potential residual confounding by weather-related factors by using natural splines of air and dew point temperatures with 3 degrees of freedom. For subgroup analyses, the investigators used information on sex, race, or ethnicity (white, nonwhite, and others), age categories (≤69, 70–74, 75–84, and ≥85 years), eligibility for Medicaid, and population density at residence (in quartiles). Subgroup-specific estimates of relative risk and absolute risk difference were obtained by fitting separate conditional logistical regression models to the data for each subgroup. To test for statistically significant differences in estimated relative risk and the absolute risk difference between categories within each subgroup (e.g., male vs female), they used a 2-sample test, based on the point estimate and standard error. They explored the health effects at lower levels of exposure by performing subanalyses with cases restricted to those occurring on days with daily air pollution concentrations below 25 $\mu g/m^3$ for $PM_{2.5}$ and 60 ppb for O_3.

Key Findings Reported by the Investigators

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

The O_3 prediction model performed similarly well, with an overall R^2 of 0.80 (Di et al. 2017b). (Note that this publication reports the R^2 as 0.76 [range 0.74 to 0.80]; presumably this is because the authors improved the model after publication of the earlier article [Di et al. 2017b].)

The average of 8-hour daily warm-season O_3 concentrations across the country during the study period ranged from 36 to 56 ppb (5th and 95th percentiles, respectively), with an average of 46.3 ppb during the study period. The investigators found a west–east gradient in the O_3 level, with the model performance being the best in the middle Atlantic, south Atlantic, east north Central, west south

Central, and the Pacific States regions. Model performance was not affected by the year, so no year-to-year trend in model fit was observed. Seasonal trends in model performance were also apparent, with the R^2 being highest in the autumn, followed by summer, spring and winter (R^2 values of 0.75, 0.71, 0.68, and 0.67, respectively). Ozone concentrations were the highest in the Mountain region and in California and were lower in the eastern states. Annual $PM_{2.5}$ and warm-season O_3 concentrations were only weakly correlated, with a Pearson correlation coefficient of 0.24.

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

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

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

elevated for both $PM_{2.5}$ and O_3, with relative risk increases (RRI) of 1.61% (95% CI, 1.48%–1.74%) and 0.58% (0.46%–0.70%), respectively. In exposure–response curves, the relative risk increase rises sharply for both pollutants at a relatively low concentration and then levels out at higher levels (see Figure 8 in the Investigators Report).

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

REVIEW PANEL EVALUATION

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

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

As stated earlier, the Phase 1 report represents a snapshot of the ambitious work undertaken by the investigators. Much work, including further development of causal methods that would properly

allow for the complexities in the design of the studies and nature of the data is currently ongoing. As a whole, this work is likely to represent an important contribution to the literature on the health impacts of air pollution on older adults in the United States. The current report represents a high-quality and thorough investigation of some of the most challenging problems in environmental health.

EXPOSURE ASSESSMENT

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

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

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

Second, because U.S. EPA monitors are located for the purpose of compliance with NAAQS, they are generally placed in the more populated, urban areas where air pollution levels are higher. Consequently, the rural areas — where population density is lower and lower pollutant concentrations are found — are not as intensively monitored, and the model may be more prone to larger error in such areas. Further, rural ZIP codes generally cover much larger areas than urban ZIP

codes. The potential impact of this on exposure estimates can be seen, for example, in the lower R^2 values for $PM_{2.5}$ estimates for the mountain region (see Di et al. 2016, Figure 1). Although only about 25% of the U.S. population lives 20 km or farther away from the nearest monitoring station — primarily in rural areas — these are the residents of potentially greatest interest, in the context of this study, because of their lower exposures to pollutants; therefore, the nature, size, and potential impact of these errors is important to understand.

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

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

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

HEALTH EFFECTS: COHORT STUDY

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

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

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

Temporal Confounding

Although the investigators have used the AG formulation of the Cox proportional hazards model to better represent time-dependent variables, the Panel's biggest concern relates to the problem of the potential for temporal confounding, with both the overall nonaccidental mortality and the $PM_{2.5}$ levels declining steadily over the period of the study, 2000 to 2012. Since this is an open cohort (new

individuals enter the cohort as they enroll for Medicare), age — which is controlled in the analyses — is not necessarily strongly correlated with calendar time. As a result, confounding can occur due to the contributions of both age and calendar time. In this study, however, there was no adjustment for calendar time and age was included in the models using five-year categories. Although the Panel understands that there are computational challenges to including a finer resolution for age, the supplementary materials accompanying the article by Di and colleagues (2017c, Supplementary Appendix) show that the hazard ratio drops from 1.07, when a five-year age category is used, to 1.05 when it is replaced with a three-year age category. This suggests that this question is unresolved and deserves more attention. Similarly, the Panel acknowledges that disentangling secular trends from any possible causal effect of $PM_{2.5}$ on mortality can be challenging and that including year in the models may over-adjust for exposure by removing true variability over time. Regardless, the inability to adequately account for potential bias due to temporal trends introduces a large element of uncertainty in interpreting the study's findings to date.

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

Potential for Residual Confounding

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

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

Spatial Differences Another issue here is the different scales at which the exposure and health models operate. The Panel has concerns about the impact of the likely exposure misclassification and confounding related to the spatial differences between aggregated summaries of exposures (1 km × 1 km) and residential locations (at the ZIP code level). The Panel appreciates that the health and

covariate data are available only as aggregated ZIP-code-level values and looks forward to the results of the planned analyses in the Final Report, in which the investigators plan to explore exposure measurement error in the health analyses using a causal-inference framework. The Panel is also aware that the exposure measurement error correction methodology for spatially varying pollutants in multipollutant research is in its infancy (e.g., Bergen et al. 2016; Szpiro and Paciorek 2013), and even more so in the causal inference framework — as duly acknowledged by the investigators — so it is not surprising that Dominici and colleagues did not yet address this in their extensive work.

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

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

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

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

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

Precision of Effect Estimates

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

Other Pollutants

Dominici and colleagues have looked at mortality associations with both $PM_{2.5}$ and O_3; this is another strength of this study. However, other pollutants may also confound the associations between $PM_{2.5}$ and O_3 with mortality. The Panel looks forward to the results of ongoing work to strengthening

the current exposure models (e.g., using data from the IMPROVE network), including a model of NO_2 and possibly PM composition.

HEALTH EFFECTS: CASE–CROSSOVER STUDY

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

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

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

SHARING OF MODELS AND DATA

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

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

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

The investigators' commitment to making their data and methods publicly available is noteworthy and welcomed, thereby enabling other investigators to access the data, to test different approaches to the analysis, and moving science forward.

Causal Inference Models

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

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

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

- It is not immediately clear whether $PM_{2.5}$ concentrations monitored inside a grid cell are error-free exposures for that grid cell, as the investigators assume. Ideally, one would use more flexible methods to allow for the possibility of such errors.
- Given that the internal validation study for the RC step is based on data from monitored locations (likely higher pollution locations compared to nonmonitored locations), it is very likely this sample will be systematically different from the main study sample. Specifically, this situation might violate some of the assumptions such as "transportability" (i.e., *relationship between true (X) and error-prone (W) exposures, conditional on covariates (D), would be the same in the validation study where X is observed and in the main study in which it is not*)**.** The

extensive simulation study does not appear to address this issue. Moreover, it doesn't grapple with the complexities of air pollution exposure, the impact of the complicated exposure modeling that produces the exposure estimates and their associated measurement error, or the complicated spatial structures of exposure, outcome, and covariates. This raises questions about the usefulness of this method in the real context of the epidemiological analyses performed in this study.

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

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

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

A common limitation of both these approaches stems from the different spatial refinement of the data, in other words, between ambient air pollution concentration estimates (at 1 km^2, which are then

aggregated to the ZIP code level) and data on health and other covariates (available at the ZIP code level). Neither of the new techniques appears to try to deal with this complexity. This continues to raise questions of exposure error and confounding that potentially affects the primary analyses, a limitation which the investigators specifically note as well. As this work proceeds, a clarification and better understanding of these issues and their impact would be important to the successful completion of the full analyses in this project.

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

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

Dominici and colleagues have conducted an extensive and innovative set of initial analyses in these extraordinarily large air pollution and health data sets. They have conducted two distinct types of analyses: a cohort-based analysis of long-term exposures and a case–crossover-based analysis of short-term exposures. They report positive associations of both $PM_{2.5}$ and O_3 with all-cause mortality, with associations extending to the lowest concentrations and with little evidence of a threshold in these initial analyses. These findings met the criteria for statistical significance, although, as noted earlier, it is important to not over-interpret the statistical robustness of results derived from such a very large data set (Meng 2018). To their credit, the investigators also conducted a range of sensitivity analyses, and they also attempted to control for many key potential confounders in their cohort study that were available in the larger data set, as well as in the smaller

Medicare Beneficiaries Survey; in all the analyses to date, these further analyses did not meaningfully change the initial findings of associations.

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

- *Further analyses of measured and unmeasured confounders*: While the investigators applied the range of data on confounders available to them and adopted the AG approach, which offered some advantage over the traditional Cox proportional hazards method in addressing some confounding due to time-dependent covariates, significant questions remain. The Panel discussed these in some detail and would like to highlight here some that will need to be further analyzed:
    - *Potential confounding by time trends*: With air pollution and death rates having declined over the course of the cohort analyses, the degree to which potential confounding of the results may have been affected by time was not adequately analyzed in these initial analyses. The investigators have acknowledged this and indicated their plan to further analyze this important question, by conducting sensitivity analyses using a newly developed causal inference approach.
    - *Potential confounding by other pollutants*: Other air pollutants may also confound the estimates of exposure and effects seen in these analyses. The investigators did test the potential influence of O_3 exposure on PM effects — and vice versa — which was an important strength of their work. In addition, they are now developing an exposure model for NO_2 that will allow adjustment for this pollutant in their final models.
    - *Analysis of spatial confounding and geographical patterns:* As the Panel noted earlier and the investigators acknowledge, the current analyses are conducted at a national level, without fully addressing potentially significant geographical variation in air pollution (both concentrations and composition) and the underlying health status (i.e., variability in $PM_{2.5}$ levels and substantial diversity in levels of obesity across different regions).

- *Spatial scales and the hybrid model:* There are several spatial scales of the many variables in both the long-term and short-term analyses, and the resulting complex quasi-ecologic (hybrid) nature of these analyses make it difficult to fully understand the implications of these. For example, as the Panel noted earlier — and despite the considerable efforts by the investigators to estimate exposure accurately — there are some potential sources of error that may affect results. These include, though may not be limited to (1) potential underestimation of rural concentration levels due to the relative paucity of ground monitors for evaluation and training in those areas; and (2) the potential differences between exposures estimated at a 1 km^2 grid but then applied to health data at the ZIP code level. Although it may not be possible to fully eliminate exposure error from an observational study such as this, the investigators will greatly enhance their final efforts by making every effort to quantify these errors and ideally to account for them in the health analyses.
- *Development, testing, and application of causal inference methods in the full population:* As noted earlier, these analyses would benefit from rigorous application of causal inference methods to the full cohort. To their credit, the investigators have taken initial steps toward developing two such methods and continuing to work on them. Properly developed and applied, these methods can also address concerns about residual confounding. The Panel has noted some important questions about these and recommends that the methods be fully evaluated and then applied.

The investigators are to be congratulated for a set of extensive and creative analyses conducted in the largest air pollution and health data base to date. While initial conclusions may be drawn from these first analyses, the Panel will wait for the planned extensive further analyses to be completed before reaching full conclusions on the air pollution and public health implications of this important research.

ACKNOWLEDGMENTS

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

REFERENCES

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

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

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

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

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

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

Di Q, Rowland S, Koutrakis P, Schwartz J. 2017b. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. J Air Waste Manag Assoc

67(1):39–52: PMID: 27332675; doi:10.1080/10962247.2016.1200159.

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

Dockery DW, Pope CA, Xu X, et al. 1993. An association between air pollution and mortality in six U.S. cities. N Engl J Med 329(24):1753–1759; doi:10.1056/NEJM199312093292401.

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

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

Hales S, Blakely T, Woodward A. 2012. Air pollution and mortality in New Zealand: Cohort study. J Epidemiol Community Health 66(5):468–473; doi:10.1136/jech.2010.112490.

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

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

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

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

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

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

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

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

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

U.S. Environmental Protection Agency (U.S. EPA). 2013. Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants. Available:

www.epa.gov/isa/integrated-science-assessment-isa-ozone-and-related-photochemical-oxidants.

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

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

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

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