SCIENCE ADVISORY BOARD
A Federal Advisory Committee to the U.S. Environmental Protection Agency

January 17, 2024

EPA-SAB-24-003

The Honorable Michael S. Regan
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Transmittal of the Science Advisory Board Report *Review of BenMAP and Benefits Methods*

Dear Administrator Regan,

Please find enclosed the final report from the Science Advisory Board (SAB). The EPA's Office of Air and Radiation requested that the SAB review the Environmental Benefits and Mapping (BenMAP) software tool and the *Technical Support Document for Estimating PM2.5- and Ozone-Attributable Health Benefits (TSD)*. In response to the EPA's request, the SAB formed the SAB BenMAP and Benefits Methods Panel with subject matter experts to conduct the review.

The SAB BenMAP and Benefits Methods Panel held three virtual meetings and met in-person on March 2 – 3, 2023, to receive a briefing on the model, discuss, and deliberate on the agency's charge questions and the Panel's integrated responses. This report conveys the consensus advice of the SAB in response to the agency's charge questions. The SAB recognizes that BenMAP and the TSD are essential tools in estimating the scale of health benefits and their associated monetary values in EPA's regulatory analysis, in particular Clean Air Act regulations reducing fine particles (PM2.5) and ground-level ozone. Thus, this report offers recommendations in the spirit of continuous improvement, especially as EPA moves toward a web-based tool for BenMAP to replace the desktop version.

Overall, the SAB concludes that the BenMAP software tool provides mapping of clean air benefits across the United States that are scientifically robust and appropriate for regulatory analyses. Nevertheless, we find that the tool can be improved in several specific ways detailed in our review. Furthermore, we note that prior versions of BenMAP have been widely used by the research and policy communities and recommend that the newly developed BenMAP web

tool be released in a manner that maintains existing functionality so that it continues to be widely usable by those communities as the prior desktop version. We would like to see future developments move towards a model of collaboration between the Agency and the broader user community.

With respect to the TSD, we would like to highlight the following recommendations as highest priority (in no particular order):

1. The SAB recommends that the EPA expand the approach described in the TSD to include studies from countries beyond the U.S. and Canada and broaden the "minimum" criteria to include studies reporting a reduced form health effect estimate of a policy change.
2. The SAB recommends that the EPA present a single probabilistic mortality estimate based on pooled risk estimates with associated uncertainty ranges rather than present multiple estimates of mortality outcomes.
3. The SAB recommends that the EPA shift to the use of scenarios for demographic information, such as the Shared Socioeconomic Pathways, rather than a single deterministic projection.
4. The SAB recommends that the EPA continue to use the 20-year segmented lag to characterize the profile of health benefits from air pollution reductions over time rather than moving to a 5-year lag structure.
5. The SAB recommends that the EPA consider approaches for estimating willingness-to-pay (WTP) accounting for the full spectrum of uncertainty with respect to causality, including levels below the "likely causal" designation.
6. The SAB recommends the EPA update the application of the value of a statistical life (VSL) to mortality risk changes to make it consistent with the current literature and internally consistent within and across regulatory impact assessments. We provide recommendations for updating the income elasticity of the VSL reflecting current best practice in the literature. We also recommend updating the underlying VSL estimate, which reflects data and analyses more than three decades old.
7. The SAB recommends that the EPA incorporate results from additional concentration-response functions beyond those that meet the criteria used in the Integrated Science Assessments (ISAs) to form a more robust pooled uncertainty estimate.
8. Regardless of whether or not the EPA updates the VSL to be more consistent with current literature, the SAB recommends that the EPA extend its current approach to quantifying sources of uncertainty in benefit measures to include a more explicit assessment of uncertainty associated with VSL including the shape of the VSL's uncertainty distribution and the transferability of VSL measures from the values in EPA's *Guidelines for Preparing Economic Analyses* to the population affected by changes in air pollution.
9. To value reductions in nonfatal health risks, the SAB recommends that EPA develop more comprehensive measures by expanding its cost of illness (COI) estimates to include a wider range of averted medical and nonmedical costs and apply estimates of WTP based on a criteria-driven review of the literature using the benefits transfer framework.

When suitable WTP estimates are not available, EPA should include proxy measures at minimum in a sensitivity analysis, rather than rely solely on COI estimates. EPA should consider the following approaches: combining health-related quality of life (HRQL) and value per statistical life (VSL) estimates; relying on monetized quality-adjusted life year (QALY) estimates that consider both HRQL and duration; and other approaches to structural preference calibration.

10. The SAB recommends that EPA include the effects of air pollution exposure on both labor productivity and human capital using a WTP approach for valuation.

With respect to the online BenMAP model, the SAB prioritizes the following recommendations:

11. The SAB recommends that EPA include access to all the data elements that it has permission to distribute and that users be able to upload their own datasets.
12. The SAB recommends that future versions allow users to apply their own health impact functions.
13. The SAB recommends that EPA work towards making the BenMAP tool "open-source" and allowing researchers to contribute to source code in a managed collaborative process.
14. The SAB recommends improvements to the guidance provided for preparing air quality surfaces and using the surfaces currently available.
15. The SAB recommends that when EPA creates a new version of BenMAP (e.g., the web version) or makes updates, EPA should run standard benchmarks to ensure that things are working as intended.
16. The SAB recommends that EPA provide a description of all model inputs used to create the model outputs through the online interface.
17. The SAB recommends all future developments pursue FAIR (findability, accessibility, interoperability, and reusability) standards and choices in data, metadata, log files, code, and all datasets.

The SAB suggests how BenMAP might provide a categorization of uncertainties beyond those attributable to the concentration-response function that are currently provided, including factors such as exposures, baseline vulnerability and population. For the latter two, our report discusses the possibility of the Agency exploring socio-economic projections that span a wide range of potential futures such as the Shared Socioeconomic Pathways (SSPs).

Ultimately, it will be important for the Agency to incorporate higher spatial resolution data to enable greater characterization of the distributional impacts of exposure changes. Finally, going forward we advocate future versions of BenMAP be developed so that it is ready to incorporate forthcoming geostationary satellite data from other federal agencies.

The SAB appreciates this opportunity to review the BenMAP model and *Technical Support Document for Estimating PM2.5- and Ozone-Attributable Health Benefits.* We look forward to the EPA's response to these recommendations.

Sincerely,

/s/

Kimberly L. Jones, Ph.D.
Chair
EPA Science Advisory Board

/s/

Drew Shindell, Ph.D.
Chair
SAB BenMAP and Benefits Methods Panel

**NOTICE**

This report has been written as part of the activities of the EPA Science Advisory Board, a public advisory committee providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use. Reports of the EPA Science Advisory Board are posted on the EPA website at https://sab.epa.gov.

The SAB is a chartered federal advisory committee, operating under the Federal Advisory Committee Act (FACA; 5 U.S. Code 10). The committee provides advice to the Administrator of the U.S. Environmental Protection Agency on the scientific and technical underpinnings of the EPA's decisions. The findings and recommendations of the Committee do not represent the views of the Agency, and this document does not represent information approved or disseminated by EPA.

**U.S. Environmental Protection Agency**
**SAB BenMAP and Benefits Methods Panel**

**CHAIR**

**Dr. Drew Shindell**, Nicholas Distinguished Professor of Earth Science, Duke Global Health Initiative, Nicholas School of the Environment, Duke University, Durham, NC

**MEMBERS**

**Dr. Enid Neptune**, Associate Professor of Medicine, Department of Medicine, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD

**Dr. Kristi Pullen-Fedinick**, Executive Director, Center for Earth, Energy, and Democracy, Minneapolis, MN

**Dr. Peter Adams**, Professor, Department of Civil and Environmental Engineering and Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA

**Dr. Joseph E. Aldy**, Professor of the Practice of Public Policy, John F. Kennedy School of Government, Harvard University, Cambridge, MA

**Mr. David Bael**, Economic Policy Analyst, Minnesota Pollution Control Agency, St. Paul, MN

**Dr. Jonathan Buonocore**, Assistant Professor, Department of Environmental Health, Boston University, Boston, MA

**Dr. Mark Dickie**, Professor of Economics, University of Central Florida, Orlando, FL

**Dr. Daven Henze**, Professor, Mechanical Engineering, University of Colorado Boulder, Boulder, CO

**Dr. Ling Jin**, Research Scientist/Engineer, Lawrence Berkeley National Laboratory, Berkeley, CA

**Dr. Michael Kleeman**, Professor of Civil and Environmental Engineering, University of California-Davis, Davis, CA

**Dr. Nicolai Kuminoff**, Professor, Economics, Arizona State University, Arizona State University & NBER, Phoenix, AZ

**Dr. Rob McConnell**, Professor, Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA

**Dr. Matthew Neidell**, Professor, Department of Health Policy and Management, Mailman School of Public Health, Columbia University, New York, NY

**Dr. Stephen Newbold**, Associate Professor, Department of Economics, University of Wyoming, Laramie, WY

**Dr. Jonathan Patz**, Vilas Distinguished Achievement Professor & John P. Holton Chair of Health and the Environment, Nelson Institute for Environmental Studies and the Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI

**Ms. Lisa Robinson**, Deputy Director, Center for Health Decision Science; Senior Research Scientist, Center for Health Decision Science & Center for Risk Analysis, Harvard T.H. Chan School of Public Health, Boston, MA

**Dr. Armistead (Ted) Russell**, Howard T. Tellepsen Chair and Regents Professor of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA

**Dr. Scott Spak**, Senior Research Fellow and Associate Professor, Department of Civil and Environmental Engineering, University of Iowa, Iowa City, IA

**Dr. Christopher Tessum**, Assistant Professor, Civil and Environmental Engineering, University of Illinois, Urbana, IL

**SCIENCE ADVISORY BOARD STAFF**

**Dr. Holly Stallworth**, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

**U.S. Environmental Protection Agency**
**Science Advisory Board**

**CHAIR**

**Dr. Kimberly L. Jones**, Associate Provost and Professor of Civil and Environmental Engineering, Department of Civil and Environmental Engineering, Howard University, Washington, DC

**MEMBERS**

**Dr. C. Marjorie Aelion**, Dean Emerita and Professor Emerita, University of Massachusetts Amherst, Amherst, MA

**Dr. David T. Allen**, Gertz Regents Professor of Chemical Engineering and Director of the Center for Energy and Environmental Resources, Department of Chemical Engineering, The University of Texas, Austin, TX

**Dr. Susan Anenberg**, Chair, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC

**Dr. Florence Anoruo**, Assistant Professor of Plant and Environmental Science and Associate Research Scientist, Department of Biological and Physical Sciences, South Carolina State University, Orangeburg, SC

**Dr. Joseph Arvai**, Director of Wrigley Institute for Environmental Studies and Dana and David Dornsife Professor of Psychology, Department of Psychology and Biological Sciences, University of Southern California, Los Angeles, CA

**Dr. Maximilian Auffhammer**, Professor, Department of Agricultural and Resource Economics, University of California, Berkeley, Berkeley, CA

**Dr. Barbara D. Beck**, Principal, Gradient, Boston, MA

**Dr. Roland Benke**, Director, Renaissance Code Development, LLC, Austin, TX

**Dr. Veronica J. Berrocal**, Professor, Department of Statistics, University of California Irvine (UCI), Irvine, CA

**Dr. Tami Bond**, Scott Presidential Chair in Energy, Environment and Health, Department of Mechanical Engineering, Colorado State University, Fort Collins, CO

**Dr. Mark Borsuk**, Professor of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC

**Dr. Sylvie M. Brouder**, Professor and Wickersham Chair of Excellence in Agricultural Research, Department of Agronomy, Purdue University, West Lafayette, IN

**Dr. Jayajit Chakraborty**, Professor and Mellichamp Chair in Racial Environmental Justice, Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, CA

**Dr. Aimin Chen**, Professor of Epidemiology, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

**Dr. Weihsueh Chiu**, Professor, Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX

**Dr. John DiGiovanni**, Professor and Coulter R. Sublett Chair in Pharmacy, Division of Pharmacology and Toxicology and Department of Nutritional Sciences, Dell Pediatric Research Institute, The University of Texas at Austin, Austin, TX

**Dr. Ryan Emanuel**, Associate Professor, Nicholas School of the Environment, Duke University, Durham, NC

**Mr. Earl W. Fordham**, Deputy Director, Office of Radiation Protection, Division of Environmental Public Health, Washington Department of Health, Richland, WA

**Dr. Rebecca Fry**, Professor, Department of Environmental Sciences and Engineering, School of Global Public Health, University of North Carolina, Chapel Hill, Chapel Hill, NC

**Dr. John Guckenheimer**, Professor and Bullis Chair of Mathematics, Emeritus, Department of Mathematics, Center for Applied Mathematics, Cornell University, Ithaca, NY

**Dr. Steven P. Hamburg**, Chief Scientist, Environmental Defense Fund, Providence, RI

**Dr. Selene Hernandez-Ruiz**, Director, Laboratory and Analytical Services Division, Water Resources Mission Area, U.S. Geological Survey, Lakewood, CO

**Dr. Elena G. Irwin**, Distinguished Professor of Food, Agricultural and Environmental Sciences in Economics and Sustainability and Faculty Director for the Sustainability Institute, Department of Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, OH

**Dr. David Keiser**, Professor, Department of Resource Economics, University of Massachusetts Amherst, Amherst, MA

**Dr. Mark W. LeChevallier**, Principal, Dr. Water Consulting, LLC, Morrison, CO

**Dr. Angela M. Leung**, Clinical Associate Professor of Medicine, Department of Medicine, Division of Endocrinology, Diabetes, and Metabolism, David Geffen School of Medicine; VA Greater Los Angeles Healthcare System, University of California Los Angeles, Los Angeles, CA

**Director Lisa Lone Fight**, Director, Science, Technology, and Research Department, MHA Nation, Three Affiliated Tribes, New Town, ND

**Dr. Lala Ma**, Associate Professor, Department of Economics, Gatton College of Business and Economics, University of Kentucky, Lexington, KY

**Dr. Enid Neptune**, Associate Professor of Medicine, Department of Medicine, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD

**Dr. Sheila Olmstead**, Dean Rusk Chair and Professor of Public Affairs, Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin, Austin, TX

**Dr. Austin Omer**, Sustainable Systems Agronomist, Crop Science Commercial, Bayer U.S., Morton, IL

**Dr. Gloria Post**, Research Scientist, Division of Science and Research, New Jersey Department of Environmental Protection, Trenton, NJ

**Dr. Amanda D. Rodewald**, Garvin Professor and Senior Director of Center for Avian Population Studies, Department of Natural Resources and the Environment, Cornell Lab of Ornithology, Cornell University, Ithaca, NY

**Dr. Jonathan M. Samet**, Professor, Departments of Epidemiology and Environmental and Occupational Health, Colorado School of Public Health, Aurora, CO

**Dr. Elizabeth A. (Lianne) Sheppard**, Rohm and Haas Professor in Public Health Sciences, Department of Environmental & Occupational Health Sciences and Department of Biostatistics, Hans Rosling Center for Population Health, University of Washington, Seattle, WA

**Dr. Drew Shindell**, Nicholas Distinguished Professor of Earth Science, Duke Global Health Initiative, Nicholas School of the Environment, Duke University, Durham, NC

**Dr. Genee Smith**, Assistant Professor, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

**Dr. Daniel O. Stram**, Professor, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA

**Dr. Peter S. Thorne**, University of Iowa Distinguished Chair and Professor, Department of Occupational & Environmental Health, College of Public Health, Director of Human Toxicology Program, University of Iowa, Iowa City, IA

**Dr. Godfrey Arinze Uzochukwu**, Senior Professor, Waste Management Institute, North Carolina Agricultural and Technical State University, Greensboro, NC

**Dr. Dominique van der Mensbrugghe**, Research Professor and Director of the Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University, West Lafayette, IN

**Dr. Wei-Hsung Wang**, Professor, Center for Energy Studies and Director of the Radiation Safety Office, Louisiana State University, Baton Rouge, LA

**Dr. June Weintraub**, Deputy Director, Center for Environmental Health, California Department of Public Health, Sacramento, CA

**Dr. Sacoby Wilson**, Professor and Director of the Center for Community Engagement, Environmental Justice, and Health (CEEJH), Maryland Institute for Applied Environmental Health, School of Public Health, University of Maryland-College Park, College Park, MD

**Dr. Douglas Wolf**, Senior Fellow, Syngenta Crop Protection LLC, Greensboro, NC

**Dr. Yiliang Zhu**, Professor, Division of Epidemiology, Biostatistics, and Preventive Medicine Department of Internal Medicine, School of Medicine, University of New Mexico, Albuquerque, NM

**SCIENCE ADVISORY BOARD STAFF**

**Dr. Thomas Armitage**, Designated Federal Officer, U.S. Environmental Protection Agency, Science Advisory Board Staff Office, Washington, DC

## TABLE OF CONTENTS

## ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| ACS | American Community Survey |
| AMI | Acute Myocardial Infarction |
| AQ | Air Quality |
| BenMAP | Benefits Mapping and Analysis Program |
| CASAC | Clean Air Scientific Advisory Committee |
| CDC | Centers for Disease Control and Prevention |
| CMAQ | Community Multi-Scale Air Quality (model) |
| COBRA | CO–Benefits Risk Assessment |
| COI | Cost of Illness |
| COVID | Coronavirus Disease |
| CPI | Consumer Price Index |
| CR | Concentration-Response (function) |
| CTM | Chemical Transport Model |
| DOT | Department of Transportation (U.S.) |
| EPA | Environmental Protection Agency (U.S.) |
| FAIR | Findability, Accessibility, Interoperability, and Reusability |
| FIPS | Federal Information Processing System |
| GEMM | Global Exposure Mortality Model |
| HHS | Department of Health and Human Services (U.S.) |
| HRQL | Health-Related Quality of Life |
| IPCC | Intergovernmental Panel on Climate Change |
| ISA | Integrated Science Assessment |
| MPAS | Model for Prediction Across Scales |
| MRADs | Minor Restricted Activity Days |
| MRRADs | Minor Restricted Respiratory Activity Days |
| NAAQS | National Ambient Air Quality Standards |
| NetCDF | Network Common Data Form |
| NBP | Nitrogen Oxides Budget Program |
| OAR | Office of Air and Radiation (EPA) |
| OECD | Organization for Economic Cooperation and Development |
| OMB | Office of Management and Budget (U.S.) |
| PDF | Probability Density Function |
| PM | Particulate Matter |
| QALY | Quality-Adjusted Life-Year |
| RIA | Regulatory Impact Analysis |
| SAB | Science Advisory Board |
| SSPs | Shared Socioeconomic Pathways |
| SVI | Social Vulnerability Index |
| TSD | Technical Support Document |
| TSP | Total Suspended Particulates |
| vQALY | Value per Quality-Adjusted Life-Year |
| VSL | Value of Statistical Life |
| WTP | Willingness to Pay |

## 1. INTRODUCTION

In the fall of 2022, the U.S. Environmental Protection Agency (EPA) Office of Air and Radiation (OAR) sought the advice of the Science Advisory Board (SAB) on its Technical Support Document (TSD) for *Estimating PM2.5- and Ozone-Attributable Health Benefits* (U.S. EPA, 2023) and the Environmental Benefits and Mapping (BenMAP) software tool (https://benmap.epa.gov). OAR developed a web version of BenMAP that calculates and monetizes the health benefits from air quality changes. The SAB was asked to peer review the new web tool which replaces a prior desktop version. OAR also provided a *User's Manual* for BenMAP (U.S. EPA, Dec. 2022) to assist with the SAB's review.

The 2023 TSD describes the process by which the Agency reviews health and economic studies in deciding how to estimate and monetize the health effects of air quality changes, the data that are used and the characterization of uncertainty. The TSD draws upon evidence published in the *Integrated Science Assessments (ISA)* for PM2.5 and ozone (U.S. EPA, 2019, 2020b).

Together, these tools enable OAR to quantify both the estimated number and the economic value of avoided air pollution-related premature deaths and illnesses as a result of changes to PM2.5 and ozone regulations. As part of its regulatory process, OAR must prepare Regulatory Impact Analyses (RIAs) that provide an estimate of the magnitude of benefits associated with lowering PM2.5 and ozone levels. Thus, OAR sought the SAB's feedback on its approach, as captured in the TSD and BenMAP tool, and presented the SAB with charge questions for its consideration. Broadly, the SAB was charged with providing feedback on OAR's approach to identifying endpoints to quantify and monetize and its calculation of health effects using BenMAP.

The SAB Staff Office formed the BenMAP and Benefits Methods Panel to conduct the peer review and brought in individuals with demonstrated expertise in the following disciplines and subject matters: air pollution epidemiology; biostatistics; risk assessment; air quality modeling; public health; data science; uncertainty analysis; and environmental economics. The Panel held its first virtual meeting on January 13, 2023, followed by a face-to-face (and hybrid) meeting on March 2–3, 2023, and two additional virtual meetings on May 16, 2023, and June 15, 2023, to deliberate integrated responses and the draft report.

The report is organized by charge question. Each section includes a charge question followed by the SAB's consensus response and recommendations. The recommendations are grouped into three tiers to indicate their priority. We offer Tier 1 recommendations as our highest priority; Tier 2 recommendations to strengthen the TSD and BenMAP model over time; and Tier 3 recommendations to be pursued further in the future, as resources allow.

## 2. RESPONSE TO CHARGE QUESTIONS

### 2.1. Charge Question 1: Health Endpoints, Epidemiological Studies and Risk Estimates

*Charge Question 1. Sections 1 and 2 of the TSD describe EPA's approach to identifying (a) health endpoints to quantify; (b) the epidemiologic studies from which we draw inferences about associations between historical changes in air quality and; (c) unit changes in risk estimated by the chosen epidemiologic studies. Please consider these sections in answering questions 1(a)-1(c).*

**2.1.a.** *Is this approach appropriate for developing a set of health endpoints, epidemiologic studies, and risk estimates for use in generating the estimated benefits associated with regulatory changes in emissions as part of the Regulatory Impact Analyses? Please explain. Please comment on the strengths and weaknesses of the approach.*

We **recommend** that the EPA continue to use the approach described in the TSD while also expanding the approach in the following ways: **(1) including studies from countries beyond the U.S. and Canada (Tier 1), (2) including impacts deemed 'suggestive' in the ISA (Tier 2), (3) including additional endpoints (Tier 2).**

Many high-quality studies exist outside of those performed in the U.S.. For example, Canadian studies are currently included in the TSD, but studies on populations from other countries are not, based on the logic that other populations or their health care systems may differ substantially from the U.S. However, Canada has national health care, as do nearly all other countries in the Organization for Economic Development and Cooperation (OECD). The differences between the U.S. and other OECD nations are likely comparable to those between the U.S. and Canada. We conclude that it would be useful to incorporate additional high-quality studies from comparable countries, when possible. We suggest the Agency prioritize expanding the literature included in their analyses in those cases in which there is a limited ability to reliably characterize the uncertainty of the exposure-response function based solely on North American studies, for example for endpoints where estimates do not exist or are few in number for the U.S. or where non-linear relationships are poorly understood. A particularly valuable result of relaxing the requirement that studies be from the U.S. or Canada is that meta-analyses and systematic reviews already available in the literature that incorporate data from both North America and elsewhere would be eligible for inclusion. A useful step would be to develop criteria on what constitutes a comparable country. A way forward might be to evaluate the distribution of key characteristics across the U.S. population, such as pollution exposures, socio-economic status, and life expectancy, and include studies from countries for which their mean values fall within the central 50% of the U.S. distribution.

The TSD likely undercounts the full range of health impacts for at least two reasons. First, many of the transient and subclinical impacts from air pollution are not included because they are not widely measured or readily monetizable. Second, many health endpoints only have a "suggestive" relationship with air pollution, where the quantitative uncertainty may be due to emerging evidence or challenges with research designs rather than a weakly established relationship.

One approach to addressing this deficiency is to consider quantifying health endpoints currently deemed "suggestive," such as birth outcomes including birth weight. Among the suggestive endpoints identified in the ISA, the EPA could prioritize those with larger monetary values and/ the greatest prevalence and severity of disease. Low birth weight, for example, is both a common and costly outcome—worthy of the analyst's attention. This prioritization would keep the additional endpoints to a smaller and manageable list. Continued reporting of individual impacts would permit only those with a 'causal' or 'likely causal' designation to be assessed when desired.

Finally, additional endpoints might be available outside the traditional literature included in the EPA's current approach to searching the literature. Research on the pollution-health relationship has become increasingly interdisciplinary, with many fields in the social sciences now conducting such research. These papers produce epidemiological-style research but are typically not published in epidemiology journals. It is important for these studies to be included in the TSD and hence we suggest the Agency's search of the literature for health endpoints and risk estimates should include social science literature, in particular economics, in addition to epidemiological literature.

As a specific example, another health endpoint worth considering is the occurrence of symptoms severe enough to lead to pharmaceutical use to alleviate them. Several studies consistently document impacts of air pollution exposure on drug use, but this is primarily in the economics (or pharmaceutical) literature since this reflects an important expenditure and thus welfare impact (Deschenes et al., 2017; Menichini and Mudu, 2010; Williams and Phanuef, 2019) rather than in the medical/epidemiological literature.

> **2.1.b.** *The TSD describes the "minimum" and "preferred" criteria used to select epidemiologic studies and subsequent risk estimates that should be used to develop Health Impact Functions for inclusion in a benefits analysis. Please comment on whether the minimum and preferred criteria identified by EPA are appropriate, on the strengths and weaknesses of the preferred criteria identified by EPA, and whether the application of these criteria has been implemented with consistency in the TSD. Do you recommend adding to or editing the criteria for the selection of studies and risk estimates, or does the SAB have recommendations for alternative criteria or changes to EPA's approach to applying those criteria? (See Section 2.2.3 and selection Excel file for the list of studies included)? Please explain.*

We conclude that the "preferred" criteria are appropriate for the selection of epidemiological studies, but that the "minimum" criteria are too narrow. We **recommend that the "minimum" criteria be broadened to include studies reporting a reduced form health effect estimate of a policy change (Tier 1). We further recommend that EPA closely follow the development of the literature with regard to differential toxicity of individual PM2.5 components (Tier 3).**

When including point estimates from specific studies, it is important to consider the many important studies do not provide a direct estimate of the marginal effect from a unit change in pollution, but instead report a reduced form estimate of a particular policy change on a health endpoint. Since many of these studies often provide a higher degree of internal validity, it would be useful to incorporate them by converting the reduced form estimates into a marginal change in pollution, commonly referred to as the Wald estimator. An example is the work of Deschenes et al. (2017), who study the

effect of the Nitrogen Oxides Budget Program (NBP) on health endpoints. This paper produced a reduced form estimate of the effect of the NBP, but also scaled the reduced form by the first stage relationship between the NBP and ozone to provide an estimate of the effect of ozone on health.

A second, closely related point is that it is important to consider incorporating studies that do not focus solely on PM2.5 but instead focus on other air pollution measures in cases for which PM2.5-based risk functions are not available. Although PM2.5 has now been monitored in the U.S. for several decades, some important contexts arise in which related pollutants, such as TSP or PM10, were instead the pollutant of interest. For example, early childhood exposure to pollution has been linked with deceased earnings in adulthood, where the gap between exposure and impacts is over 40 years (Isen et al., 2017; Voorheis 2017; Colmer and Voorheis, 2020). Another scenario arises where investigators use an historical event as part of an empirical strategy to obtain causal estimates of the effect from pollution exposure, such as the passage of the Clean Air Act Amendments or an important macro-economic event (Sanders, 2012). Alternatively, TSP or PM10 may be appropriate measures when the investigation is in a country where PM2.5 measures are only more recent, which is the case in many European nations (see charge question 1.a for the recommendation for including such countries). For example, Holub et al. (2020) study PM10 in Spain and Roth studies PM10 in London (2018) but neither incorporate PM2.5 data.

If these studies provide estimates for health or non-health endpoints that are not captured elsewhere, it would be prudent to consider including them in the EPA's approach by converting the pollutants into PM2.5 using relationships between the two. For example, Brook et al. (2011) find that PM2.5 accounts for 49% of PM10 and PM10 accounts for 44% of TSP. However, there was considerable variability among sites in the Brook et al. study, with the mean PM2.5 to PM10 ratio ranging from 0.36 to 0.65. Therefore model results for the spatially resolved ratio of PM2.5 to PM10 would likely need to be used for this conversion, which would be challenging as the relationship between TSP or PM10 and PM2.5 will vary temporally as well as spatially.

When additional outcomes are considered for which data are from an historical event or non-U.S. context, there is a potential for double counting if such outcomes are also included in endpoints currently addressed . This can be minimized by focusing on non-traditional outcomes such as productivity or human capital (discussed in charge question 4.g). Note that when the outcomes under investigation stem from latent effects, the risk of double counting is essentially non-existent; these latent effects would not be captured elsewhere.

With respect to the chemical composition of PM2.5, the EPA's health evaluation does not consider different components based on the conclusions of the ISA. Emerging evidence suggests that different sources of PM may have different toxicities (e.g., Thurston and Bell, 2021; Wang et al. 2022), thus suggesting a future consideration for the EPA to include. This may also help when incorporating evidence from outside the U.S., where PM composition might be substantially different. The EPA should follow developments in this area with an eye toward eventually including differential toxicities of various PM components should adequate data become available.

**2.1.c.** *Considering existing OMB and EPA guidance on preparing economic analyses, please comment on: (1) the conditions under which the science or other best practice guidelines would*

*support EPA presenting multiple estimates of pollutant-attributable mortality; (2) whether the mortality estimates described in the TSD capture the range of potential benefits of PM2.5 and ozone reduction; and (3) the scientific and analytic advantages and disadvantages of reporting benefits calculated using a pooled risk estimate versus reporting benefits calculated using the single estimate best characterizing risk.*

**We recommend that the EPA present a single probabilistic mortality estimate based on pooled risk estimates with associated uncertainty ranges rather than present multiple estimates of mortality outcomes (Tier 1). This will require development of a systematic method for pooling the available data that weights the studies to account for differences in credibility and provides risk estimate uncertainties (Tier 1).**

Guidance documents developed by both the Office of Management and Budget (OMB) and EPA encourage examining uncertainty broadly construed (OMB, 2003; U.S. EPA, 2014b).[1] Risk estimates from the top epidemiologic studies sometimes differ by a factor of two or more. Presenting multiple estimates drawn directly from the primary literature is one way to convey the prevailing uncertainty and seems consistent with both sets of guidelines. The mortality estimates from PM2.5 and ozone reduction described in the TSD are presented in this spirit, and we conclude that these raw estimates provide a reasonable characterization of the range of potential benefits of PM2.5 and ozone based on current understanding.

Yet, beyond merely presenting two or more alternative estimates of population risk parameters, we believe that a preferred approach would be to develop a single synthetic estimate and associated characterization of uncertainty that combines information from all relevant primary studies. The results of such a synthesis could take the form of a central point estimate with an associated range for sensitivity analysis or a probability distribution for Monte Carlo analysis. We encourage the Agency to examine whether one or more standard methods of meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2021; Cooper, Hedges, & Valentine, 2019; Burnett et al., 2018, Stanley and Doucouliagos 2011) or other forms of information synthesis (Manski CF, 2020; Manski, 2010) could be used for this purpose, and to support internal and external research and development of modified or new methods as needed. Presentation of results using individual risk functions should still be maintained to allow individual users flexibility even if a pooled risk became the primary risk function used in benefits analyses.

A particular challenge that the EPA will face when synthesizing risk estimates will be to account for relevant heterogeneities within a single study among the exposed individuals that comprise the samples in each primary study. There are also substantial heterogeneities between studies, and the health effects of pollutant exposures are known to be moderated by age, sex, socio-economic status, and other factors, leading to heterogeneities since these factors will vary across studies in observable and unobservable ways. For example, the EPA may need to combine estimates of the effect of PM2.5 concentrations on mortality risks from one study that examined individuals aged 40 and above from middle- and high-income households with an estimate from another study that examined individuals

[1] OMB Circular A-4 was updated on November 9, 2023, after the Panel completed this report. We understand that EPA will be taking into account the effect of any updates on the recommendations in this report. The updated Circular may be found at https://www.whitehouse.gov/wp-content/uploads/2023/11/CircularA-4.pdf.

aged 50 and below from low- and middle-income households. Some form of structural or model-based meta-analysis may be needed to account for these and other moderating factors (Jak, 2015; Jak & Cheung, 2020; Jak, Li, Kolbe, de Jonge, & Cheung, 2021). In **Appendix A: Pooling Risk Estimates**, we provide a simplified example to show how risk estimates pertaining to distinct or overlapping age ranges can be synthesized in an internally consistent manner. Burnett et al. (2018) provides another example of a method to pool results from different age groups. These examples consider only one dimension of heterogeneity, whereas in many cases analysts will need to account for multiple heterogeneities to provide a credible structural equation for generalizing risk estimates to a broader population or transporting estimates to new policy settings (Tipton & Hartman, 2023), so the example illustrates both the opportunities and the challenges of structural meta-analysis.

A separate challenge when pooling evidence involves the assignment of weights on individual estimates from different studies. Conventional methods of meta-analysis pool estimates using weights that depend on the estimated sampling variability of each primary estimate (standard errors) and the overall heterogeneity among the estimates (Kepes, Wang, & Cortina, 2023). However, outside of the data collection phase when choices to include or exclude the available estimates are made, conventional meta-analytic methods typically do not account for differences in aspects of the identification strategies and other study design features that may affect the internal and external validity of the resulting estimates. A more efficient study with low standard errors might be based on a design that is less likely to isolate a generalizable or transportable causal treatment effect than another study with higher standard errors, which would set up a trade-off between precision and accuracy when combining evidence from the two studies. Therefore, we encourage the EPA to explore methods for adjusting the weights it assigns to pooled risk estimates to reflect the Agency's considered judgments about differences in validity among the primary studies. Some previous meta-analyses have used a form of quality weighting (Ahn and Becker, 2011; Valentine, 2019), but we are aware of none that have applied the approach we propose in **Appendix B: Validity Weighting in Meta-Analysis**.

We do not have a silver-bullet recommendation on this issue. Different methods of characterizing uncertainty and synthesizing risk estimates will have competing advantages and disadvantages. At one extreme, merely presenting multiple primary risk estimates may be the simplest way to convey some of the prevailing scientific uncertainty, yet it places a higher burden of interpretation and synthesis on the decision-makers who must make use of the estimates. At the other extreme, using some form of structural meta-analysis would allow analysts with the relevant subject matter and statistical expertise to make the judgments and perform the probability calculations that are notoriously difficult for most people (Benjamin, 2019), and it would give decision-makers a more directly actionable bottom-line result. But this approach would be more challenging for analysts to implement and would require making a series of simplifying assumptions and explaining these in a manner that is transparent and readily comprehensible to most readers. Of course, this is not a binary choice. The EPA could present multiple estimates in their raw form *and* present the results of a meta-analysis that produced a synthesized point estimate and uncertainty range or distribution, with full explanation of the intermediate and final results, including the key assumptions that are needed to justify each step in the progression.

**2.1.d.** *Currently, EPA updates its selection of health studies approximately every five years, in concert with each new ISA. What is the appropriate frequency for EPA to update its selection of health studies? What would be too frequent? What would be insufficiently frequent?*

**We recommend that the EPA continue to update its selection of studies at five-year intervals along with the ISA, and that it consider providing interim updates at its discretion when substantial advances in the science are reported (Tier 3)**.

The ideal periodicity of review would be as frequently as the science demands. Since the science is rapidly emerging at a rate that challenges EPA's 5-year periodicity, we suggest that the EPA consider creation of a supplemental procedure that would allow for more rapid updating when new scientific evidence emerges. The details of such a procedure appears to be more of a practical question for EPA than a question that outside scientists could answer. The SAB is mindful that EPA is tethered to a 5-year review schedule under the Clean Air Act[2], and the review concluded that updating less frequently than at 5-year intervals would be insufficient.

Given that EPA is constrained by law and its *Integrated Science Assessments* development process, we suggest a secondary step for the EPA to facilitate the process of outside researchers uploading their own beta coefficients when a new study becomes available. Naturally the latter would allow for any concentration-response (CR) function but would not necessarily be "officially endorsed" by EPA.

### 2.2. Charge Question 2: Baseline Rates of Death and Disease

*Charge Question 2. Sections 3 and 6.3 of the TSD describe the baseline rates of death and disease used to quantify counts of pollutant-attributable effects. Please consider the approaches used to generate these baselines as you answer questions 2(a)-2(e).*

**2.2.a.** *Are these rates appropriately matched to each outcome according to ICD-9 or ICD-10 code?*

The SAB endorses the current approach and concludes that the ICD-9 or ICD-10 codes appear to be appropriately matched to each outcome.

**2.2.b***. Are these rates resolved to the appropriate spatial and temporal scale for performing benefits analyses?*

**We recommend that all baseline rates of death and disease be resolved at the county level, as some already are, or as close as possible to that level, and that the EPA balance the use of the most up-to-date rates available with the stability of longer-term data (Tier 2).**

Regarding spatial scales, the EPA provides information on the geographical distribution of some background disease rates in the TSD. Mortality is stratified to age, cause, race, and county, which is adequate. All-cause mortality rates are specified at the census tract level, which is even more useful than the county level. While the panel recognizes that these are unlikely to be available for all specific individual health endpoints of interest, it encourages the Agency to continue pursuing higher resolution baseline rates of death and disease, especially as it moves towards more systematic and detailed distributional analyses. We encourage the Agency to expand baseline health (and population)

[2] In Section 42 U.S.C. §7409(d)(2), the Clean Air Act directs the Clean Air Scientific Advisory Committee (CASAC) to review the National Ambient Air Quality Standards (NAAQS) "at five-year intervals" and recommend to the EPA Administrator "any new national ambient air quality standards and revisions …as may be appropriate."

datasets to better represent indigenous communities at a locality level. Other data are at varying degrees of geographical detail and other forms of stratification, but the information provided in the TSD is inadequate to evaluate the level of geographic detail underlying the analysis for some of the morbidity outcomes. The possibilities remaining with the documentation in the TSD imply very different levels of geographic specificity, so it is important that the TSD be clear about the geographic specificity of data coverage. It would be helpful for the TSD to provide a map of the detail of data coverage, and quantify the populations covered by each level of geographic specificity.

The EPA should strive toward having as much of this data as possible at the county level, and to evaluate the adequacy of the data used in the TSD, the EPA should provide information on how much data are at what level of geographic detail. Currently, the TSD lists the following health outcomes as using national, regional, state, and/or county as levels of detail: Hospitalizations, emergency department visits, nonfatal acute myocardial infarction, and Alzheimer's disease. However, there is no information on coverage for each of these levels of geographic specificity. For example, with information provided in the TSD it could be that 37 states have county-level information, 12 have state-level, and 1 state is using a nationwide average; alternatively, it could also be that 1 state had county-level data, 3 had a state average, and 46 were using a nationwide average.

The following health outcomes appear to use national averages: asthma onset, Parkinson's disease, allergic rhinitis, cardiac arrest, lung cancer, stroke, work loss days, school loss days, and minor restricted activity days. This is inadequate as it misses hotspots in background disease, which have been noted for asthma (New York City Health Department Environment and Health Data Portal), and likely exist for other health impacts.

The TSD's background rates of death and disease could benefit from an assessment of the data quality in addition to the questions of coverage and resolution. Such an assessment could focus on the following questions:

- Is there comprehensive nationwide coverage of background rates for a disease? Or are there some areas of the country missing?
- What is the core geographical resolution of the data being used? Are individual counties each reporting individual rates for each county or is a state or national average being used?

We also suggest that the EPA explore the use of alternative datasets and methods that exist that could be incorporated into the TSD, BenMAP and future RIAs, specifically:

- data available from Medicare and Medicaid claims datasets (Di et al., 2017a; Di et al., 2017b; Wei et al., 2021; Danesh et al., 2021; Eum et al., 2019);
- data available from state, regional, and city departments of public health, U.S. Centers for Disease Control, or similar agencies (e.g., New York City Health Department; Center for Disease Control); and
- imputation, interpolation, and downscaling methods, along with synthetic population methods for deriving more geographically resolved background rates of health outcomes (Levy et al., 2014).

However, this could be challenging given inconsistencies in existing data.

Regarding timescale, there is a balancing act between finding a long-term, robust estimate of background health rates and being up-to-date. This is made even more challenging with COVID likely producing a categorical change in background vulnerability of the population and changing disease rates. EPA should endeavor to balance the most recent years with data available that apply most readily to years being examined with use of an adequate number of years of data to provide a stable estimate of background rates, which likely depends upon the particular health outcome.

It would also be appropriate for EPA to incorporate background disease rates in a formal uncertainty analysis, including variation and general uncertainty, along with uncertainty introduced by having data at low-resolution.

**2.2.c.** *Are the assumptions EPA makes in projecting death rates reasonable?*

The SAB concludes that the assumptions EPA makes are reasonable but **recommends that the EPA move toward using projections that incorporate uncertainty in projected death rates such as the Shared Socioeconomic Pathways' projections (Tier 2).**

The assumptions made in the TSD to project death rates are reasonable to the extent that it is reasonable or necessary to make a single deterministic prediction of future death rates. However, there exists substantial uncertainty regarding the future development of factors that influence death rates, including climate change, national and international politics, and technological development between now and 2060. Because of this uncertainty, a single deterministic prediction of related phenomena is unlikely to be accurate. Scenario analysis—making predictions under a range of assumed scenarios meant to capture the range of plausible outcomes—is often considered more appropriate than deterministic predictions when reasoning about the future under high uncertainty (Duinker and Greig, 2007). The "Shared Socioeconomic Pathways" (SSPs; Riahi et al., 2017) framework has recently gained widespread adoption in the research community as a shared framework for energy and environmental analysis for an uncertain future and could be useful in the context of EPA's benefits analyses. The EPA should explore the possibility of using SSP-based projections for death rates, for example based on the work of Hauer (2019).

By default, the SSPs are specified at a spatial resolution much coarser than is practical for use in the context of air pollution health impact assessment. SSP information has been downscaled to regional and local scales in several studies on a case-by-case basis, but there do not yet exist emergent best practices for such downscaling (O'Neill et al., 2020). Downscaling has been used in several relevant contexts, including air pollution, population, and temperature data (Fabian et al., 2014; Limaye et al., 2018; Petkova et al., 2013; Di et al., 2019; Voorhees et al., 2011; Sun et al., 2022). The methods used include interpolation based on other available data, regression approaches, and machine learning. We also encourage the EPA to update its own Integrated Climate and Land-Use Scenarios (ICLUS) project[3] to incorporate more recent data within its projections of demographic trends. This project has produced spatially explicit projections of population and land-use for the conterminous United States on an annual basis through 2100 for a number of SSPs. These include domestic migration associated

[3] EPA's Integrated Climate and Land-Use Scenarios (ICLUS) project provides downloads of land use and population projections at https://www.epa.gov/gcx/about-iclus.

with climate change, as well as other factors, and hence updated versions of these could provide valuable information for high-resolution death rate projections.

Finally, we encourage the Agency to work towards aligning projections of death rates used for air quality-related benefits analyses and those used in the analyses of the social cost of greenhouses gases. This applies not only to death rates, but also to morbidity rates and population characteristics discussed in response to later charge question subsections.

**2.2.d.** *Is the approach used to estimate race- and ethnicity-stratified mortality incidence appropriate?*

The SAB concludes that the current approach is appropriate but, as a longer-term goal, we **recommend that the EPA look to include sub-county information and additional datasets related to the impact of COVID on mortality incidence (Tier 3).**

The EPA should continuously search for alternative datasets and methods to develop baseline mortality and morbidity rate data that is stratified by race and ethnicity, while also being mindful of existing constraints due to privacy concerns and suppressed data. The method the EPA applied in the TSD is adequate to develop mortality estimates for county-level estimates, pre-COVID, but this may not adequately represent the impact of COVID, or hotspots in mortality that exist at the sub-county level. We conclude that, as in our response to question 2b, it would be appropriate to consider using data from Medicare, Medicaid, datasets maintained by other federal, state, regional, and city health agencies, or imputation methods in order to develop mortality estimates below county level.

According to the TSD, race stratification used 2007 – 2016 as the date range whereas the 2012 – 2014 timeframe was used for the baseline. It was unclear why different date ranges would be used and it would be helpful for the Agency to expand upon the use of those specific dates.

It would also be helpful to understand, or at least acknowledge, the impact that the COVID pandemic will have on both absolute death rate and mortality attributable to specific causes. We note that there may be an impact of COVID on overall lung and immune functioning, and that the virus can exacerbate death rates associated with causes that are not explicitly denoted as COVID-related.

**2.2.e.** *Should EPA also consider projecting all or some morbidity rates?*

The SAB **recommends that the EPA explore the availability of projections that reflect differences in baseline morbidity rates across race/ethnicity, and across geographies, such as those created under the Shared Socioeconomic Pathways initiative (Tier 3).**

Projecting morbidity data is challenging. Consider, for example, changes in rates of cardiovascular disease, which declined steadily for several decades throughout the latter 20th century (Kones and Rumana, 2017; Young et al., 2010). Until recently, it would have been reasonable to assume that the decline in some risk factors – hypertension, high cholesterol, smoking -- and in cardiovascular disease itself would have continued to decline. However, there was a deceleration in the decline in cardiovascular disease mortality in the last decade (Reither et al., 2011; Avanth et al., 2023)). The

epidemic of obesity and associated type 2 diabetes has also been projected to increase morbidity and mortality, especially among youth and young adults as they age (Stephen et al., 2016). Even before the COVID pandemic stagnation in the increase in life expectancy occurred in some US demographic groups. More recently there are promising new pharmacological treatments for obesity and associated metabolic disease (Akoumianakis et al., 2023). The impact of these treatments on long-term morbidity and mortality is still under study, so there is uncertainty in projecting changes into the future. In addition, some states have limited data even on current county-level morbidity rates, as described in Section 6.3 of the TSD.

Given these uncertainties, for some outcomes using current morbidity rates may be a more defensible option than projecting changes in morbidity into the future. Nevertheless, one suggestion for how to project morbidity rates is the Shared Socioeconomic Pathways (SSP) developed by the research community described in greater detail in charge question 2c. Predictions can be made under a range of assumed scenarios meant to capture the range of plausible outcomes. Evaluating the effects of policies under these different scenarios could be part of sensitivity tests and uncertainty analysis. Scenarios might evaluate potential changes in full population morbidity rates due to trends like diet, medication, and COVID, along with equity/inequity of changes across race/ethnicities, income, educational attainment, and other demographic factors. Estimating these changes in covariates into the future and their impact on morbidity will be labor intensive, so it will be important to consider carefully which changes are likely and are also likely to have substantial impact on morbidity, and to focus on them. Regardless of the choice around one set of estimates or a set of scenarios, the EPA should develop projections of morbidity rates that reflect differences in baseline morbidity rates across race/ethnicity, and across geographies, in as much geographical detail as possible.

Finally, appropriate qualitative consideration of uncertainty of current morbidity rates and of current population trends in rates, is appropriate in interpreting the pollutant-attributable effect estimates, including those from BenMAP. If BenMAP impact assessment continues to be updated in 5-year intervals, this provides an opportunity for corresponding updates in local morbidity rates. The trend in the age distribution of the population is relatively predictable at a local level, notwithstanding uncertainties due to factors such as migration, and this is likely to have a larger impact on health costs and benefits of PM2.5 and ozone than less certain trends in changing morbidity not related to an aging population.

### 2.3. Charge Question 3: Population and Demographic Data

*Charge Question 3: Section 4 describes the population and demographic data used for quantifying health effects. Please reference this section in answering questions 3(a)-3(d).*

**2.3.a.** *Does the demographic information used to quantify pollutant-attributable cases as described in Section 4 reflect the best available science?*

The SAB **recommends that the EPA shift to the use of scenarios for demographic information, such as the Shared Socioeconomic Pathways, rather than a single deterministic projection (Tier 1).** We also **recommend that the Agency pursue further intra-urban level spatial resolution, extending analyses to neighboring nations, and developing models that account for the interdependence of air pollution changes themselves on population demographics (Tier 3).**

Based on the TSD and clarifications from EPA staff, our understanding is that demographic data are determined by starting with the most recent available decennial Census data (for year 2010 currently) and then projecting it to years of interest using proprietary data from Woods & Poole. It appears that all demographic data are used at the Census-tract level of spatial aggregation. As discussed elsewhere in this report (for example, in response to charge question 2), scenario analysis is a more appropriate way to analyze future population, rather than making deterministic predictions. The use of the SSPs for this purpose has gained wide adoption, and SSP-based population projections are available, for example by Hauer (2019). These data are freely available, transparently documented, and generated by the academic community, which is preferable to the closed, inscrutable nature of the Woods & Poole projections. In addition, the academic community has developed SSP-specific projections for GDP per capita that could be used to consistently adjust WTP measures for future growth in real income, as outlined in TSD section 5.4. Dellink et al. (2017) and Wang and Sun (2022) provide examples.

The use of census-tract level data will provide fairly high spatial resolution for demographic information. We encourage the Agency to push as much as practical toward utilizing intra-urban spatial resolution in their health impact assessments, including use of demographic data at block or block-group resolution, as concentration levels typically co-vary with demographic characteristics within urban areas (Paolella et al., 2018). This effort is sometimes hindered by data suppression at high resolution, but this can be addressed in many cases by using data imputation (Levy et al., 2014). However, depending on why data are missing, imputation can sometimes bias results and create problems for estimation and inference, so that use of imputation should be carefully considered (e.g., Lillard, Smith, and Welch 1986; Hirsch and Schumacher 2004; Bollinger and Hirsch 2006; Little and Rubin 2014).

Additionally, we note that current EPA methods only consider people within the borders of the United States. However, residents of other countries (in particular, Canada and Mexico) are also exposed to pollution originating in the United States. In light of a recent return by the EPA to the long-running practice of considering impacts to populations outside of the U.S. in its analysis of climate change impacts (Interagency Working Group Paper, 2021), we encourage the Agency to also consider doing this in its air quality impact assessments.

Finally, we would like to highlight that some of the exogenous assumptions that go into population projections are in fact partially endogenous to any air quality scenario that is being analyzed. For example, reductions in air pollution will decrease the rate of death, leading to increased population. Similarly, if air pollution improves in a specific area, housing costs may rise, attracting more affluent people while displacing the previous residents (Patterson and Harley, 2019). We encourage the Agency to move toward approaches to modeling that can capture this type of dynamical interdependency. This may require interdisciplinary collaboration within EPA, or the adoption of more flexible software development frameworks and techniques to promote coupling among different models and is likely to require substantial effort.

**2.3.b**. *The Woods & Poole projected population is used to quantify and value PM2.5 and ozone health impacts. Is the Woods & Poole data appropriate for this purpose? Are there alternative*

*projections that EPA should consider using instead? Why might your proposed alterative be preferable?*

The SAB **recommends that the EPA shift to the use of publicly available, peer-reviewed scenarios for demographic information, such as the Shared Socioeconomic Pathways, rather than the proprietary, single deterministic projection of Woods & Poole (Tier 1).**

There are a number of datasets providing population projections outside of Woods & Poole that may be appropriate for the purpose. Most notably, there are the population projections underlying the SSPs that are used by the Intergovernmental Panel on Climate Change (IPCC) (Hauer 2019). Those datasets are open access and publicly available (Hauer 2018). In addition to that dataset being fully open access (methods and code available publicly), publicly available, and peer-reviewed, the dataset provides five alternative projections for population throughout the United States, giving EPA the capacity to evaluate uncertainty and sensitivity of results to alternative population projections, something we have stressed in previous charge questions. We also encourage the Agency to consider updating its own ICLUS project projections with more recent data that could provide another option for population projections consistent with the SSPs.

**2.3.c**. *Are the age, sex, race and ethnicity strata in the population data used to quantify effects appropriately delineated and appropriate for quantifying effects across subpopulations?*

While the current data are appropriate, the SAB **recommends that EPA consider adding additional stratification to include multi-racial people, sexual minorities, and individuals with co-morbidities (Tier 2).**

The TSD does not include extensive information regarding which strata are used, stating only that projections are "stratified by race/ethnicity, age, and sex". In the BenMAP *User's Manual*, Table D-8 states that the race strata are White, Black, Native American, and Asian. Age strata appear to be < 1 year, 1-4 years, 5-14 years, 15-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65-74 years, 75-84 years, and 85+ years for race- or ethnicity-stratified data (Appendix D). It appears that there are also ethnicity strata (Hispanic and non-Hispanic; Table J-5). The gender strata are apparently Male and Female (Table J-5).

In addition to the categories above, we recommend that the Agency consider additionally quantifying impacts to multi-racial people, sexual minorities, and individuals with co-morbidities. These groups are all known to have unique susceptibilities to various exposures (Goldsmith and Bell, 2022). We recognize that these efforts may be hindered by data limitations but encourage the Agency to explore the use of data imputation to fill any gaps, motivated by the consideration that an imperfect quantification of an important issue is often better than no quantification at all. A potential example of imputation is that described in Levy et al. (2014), though as noted in our response to 2.3a there are challenges associated with imputation that need to be considered.

We would also like to stress the importance of the ability to analyze impacts stratified by demographic group, and to encourage the Agency to fully support the use of BenMAP-CE until these capabilities have been fully integrated into the new web version of the model.

**2.3.d**. *Are the data used to estimate the number of individuals according to poverty status, educational attainment, and linguistic isolation appropriate?*

The SAB concludes that the data used are appropriate but we r**ecommend that the EPA provide additional documentation on these data (Tier 1) and explore the use of additional indicators of socio-economic status (Tier 3).**

Though the 2023 TSD did not provide specific information on the methods used for poverty status, educational attainment and linguistic isolation, details are given in Appendix J of the BenMAP *User's Manual*. These descriptions were helpful and generally represented reasonable approaches. There were several areas that lacked clarity including the specific variables from the American Community Survey (ACS) (U.S. Bureau of Labor Statistics, 2015) which were used in defining each of these areas. Though narrative descriptions were provided (e.g., "below poverty line"), it would be helpful for the Agency to list the specific variable used (e.g., B17001_002E) to aid in evaluation of the appropriateness of the variable.

Similarly, specificity was not given about the specific variable used when multiple options were available – e.g., for linguistic isolation, two options - "less than very well" and "less than well" - were both defined as options. To increase transparency and the enhance the ability of external review of the methodologies, the variables used in the model should be explicitly defined. Additionally, if multiple variables are available for the same outcome, sensitivity analysis to determine the impacts of one option over another would be appropriate.

We also note that the mechanism for selecting the socio-economic factors to include in EPA's analysis is not clear, and that the Agency's approach utilizes a smaller set of indicators relative to indices such as the Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI) (Center for Disease Control, 2018). In the CDC SVI, e.g., education and poverty are combined in a single variable defining socioeconomic status. Additional variables for socioeconomic status include income and unemployment variables. Sensitivity testing should be performed to determine the impacts of using a more inclusive definition of socioeconomic status versus poverty and education in isolation. The Agency should, at a minimum, evaluate the impacts of additional variables (e.g., those from the CDC SVI) on modeled outcomes in BenMAP.

Finally, we note that the BenMAP demographic data are relatively coarse and lack information on household composition and housing characteristics that can make populations more vulnerable to the impacts of air pollution and other forms of environmental degradation. A more robust approach to identifying susceptible (and vulnerable) populations should be utilized whenever possible, to ensure that the Agency most accurately calculates both the burden and benefits associated with specific regulatory actions. The CDC SVI data are available at the tract level, which could make for a relatively easy integration into the BenMAP model.

**2.3.e**. *Are there other characteristics across which EPA should report stratified results? If so, what characteristics, and what would be appropriate ways to measure those characteristics?*

The SAB concludes that the characterization of stratified results is acceptable but we **suggest additional characteristics that could be added (Tier 3).**

The characteristics included in the TSD (age, sex, race-ethnicity and location) are acceptable. However, as discussed in prior sections, mixed race status should ultimately be developed as a characteristic and described by the component ethnicities. These are imperfectly included in the informing instruments which creates a need to examine emerging datasets that reflect this aspect of the current racial landscape of the country. Other characteristics that should ideally be captured and stratified are rates of premature birth, proximity to health care providers, and migration patterns.

### 2.4. Charge Question 4: Valuing Effects

*Charge Question 4: Sections 5 and 6.4 describe the approaches EPA uses for valuing effects. Please reference this section in answering questions 4(a)-4(j).*

*Cross-cutting Discussion of Valuation Approaches*

The SAB provides Tier 1 recommendations concerning valuation of nonfatal health risks that pertain to several charge question subparts. **We recommend that the EPA develop more comprehensive cost of illness (COI) estimates that include a wider range of averted medical and nonmedical costs and apply estimates of WTP based on a criteria-driven review of the literature using the benefits transfer framework. When suitable WTP estimates are not available, EPA should include proxy measures at least in sensitivity analysis, rather than relying solely on COI estimates. Approaches EPA should consider include combining health-related quality of life (HRQL) and value per statistical life (VSL) estimates, relying on monetized quality-adjusted life year (QALY) estimates that consider both HRQL and duration, and other approaches to structural preference calibration.**

Five subparts of this charge question (4.b, 4.d, 4.e, 4.f, 4.h) relate to the use of cost of illness (COI) estimates for valuation. To minimize repetition, we provide a cross-cutting overview of related issues and recommendations here. More details on the approach for each endpoint are then discussed in the responses to the individual questions below.

As indicated in the EPA *Guidelines for Preparing Economic Analyses* (U.S. EPA, 2014b) and other best practice guidance, e.g., from the U.S. Office of Management and Budget and U.S. Department of Health and Human Services (HHS) (OMB 2003, U.S. HHS 2016), WTP is generally the most appropriate approach for valuing reductions in fatal and nonfatal risks in benefit-cost analysis. However, for nonfatal health outcomes, at times EPA and others use COI estimates as rough proxies when WTP estimates are not available. These costs typically include direct medical costs and indirect costs associated with time losses due to illness. These time losses in turn may include changes in nonmarket (unpaid) and market (paid) time use, including both labor and leisure. Other averted costs may also be included if relevant, such as those associated with remedial education if pollution controls would decrease the risk of cognitive impairments.

Comprehensive COI estimates should include costs borne by the affected individual and costs borne by other members of society. Even if individual-level WTP estimates are available, some social costs also should be included in the benefits estimates. WTP reflects individuals' willingness to exchange their

own income for changes in their own risk. As discussed in the U.S. HHS *Guidelines for Regulatory Impact Analysis* (U.S. HHS, 2016) and elsewhere, real opportunity costs not borne by the individual (e.g., insured medical costs and caregiver time) should be added to estimates of WTP—or to other proxies for WTP (as discussed below)—to provide a comprehensive measure of value. However, costs that otherwise would be incurred by the affected individual should be excluded to avoid double-counting when WTP estimates are used. Note that out-of-pocket costs borne by the individual are likely to significantly understate that individual's WTP, since they exclude pain and suffering and other nonpecuniary effects; the extent of understatement is uncertain but may be significant.

Another challenge in COI estimation arises because many of the epidemiological studies use a particular event, such as hospital admissions, as the outcome measure. These events may relate to exacerbation of an existing case or to incidence of a new case of illness. Because the value of averting these exacerbations and new cases likely exceeds the value of averting the event itself, we recommend that EPA review analyses of administrative records and perhaps the disease modeling literature to determine whether it is possible to predict the likelihood that extent signifies worsening of existing cases or incidence of new cases. The latter literature (e.g. Drummond et al., Editors, 2012) often uses microsimulation models to predict the incidence and progression of disease.

For more information on best practices for estimating medical costs in regulatory analysis, see the 2016 and 2017a guidance developed for the U.S. HHS as well as Onukwugha et al. (2016). As recognized in the TSD, one challenge in estimating medical costs is the need to ensure that the estimates reflect opportunity costs, which may differ from posted prices. Another challenge is ensuring that the estimates are reasonably comprehensive, including in-patient and out-patient costs as well as time losses; for many endpoints, EPA only includes a subset of these costs.

Both HHS (2016, 2017b) and EPA (2020b) have developed guidance indicating best practices for valuing changes in time use. The guidance issued by each agency is similar in many respects but includes some important differences. However, the TSD does not appear to apply a consistent approach for valuing changes in time use across health outcomes.

Due to the limitations of COI estimates, we recommend that the Agency develop a strategy to reduce its dependence on COI estimates as proxies for WTP. In the near-term, EPA should conduct a criteria-driven review of the WTP literature (following a benefit transfer framework, e.g., U.S. EPA 2014, U.S. HHS 2016) to determine whether available estimates could provide a reasonable alternative or supplement to the COI measures, either as part of the primary benefits estimates or in sensitivity analysis. Several previous reviews may provide a useful starting point, including Alberini et al. (2010), Hunt and Ferguson (2010), Gerking and Dickie (2013), Hunt et al. (2016), and the European Chemicals Agency (2016). In addition, a new Organization for Economic Cooperation and Development (OECD) study (Organization for Economic Cooperation and Development, 2023) provides estimates for some health effects of concern, which should be included in this review.

When suitable estimates of WTP are not available, EPA should include proxy measures at least in sensitivity analysis rather than relying on COI estimates alone. The SAB has identified three approaches that the EPA could use to improve its ability to value nonfatal health effects, based on (1) health-related quality of life (HRQL) indices, (2) quality-adjusted life year (QALY) estimates that combine

estimates of HRQL and duration, and (3) structural benefits transfer (or "preference calibration"). A quality-adjusted life-year (QALY) is a nonmonetary measure that reflects the duration and severity of illness in a single scalar (see IOM 2006, HHS 2016, Neumann et al. 2016 for more discussion). QALYs are computed as the product of the duration of a health state and a measure of the health-related quality of life (HRQL) of the health state. The HRQL is an index usually measured on a scale of zero (as bad as death) to one (full health). QALYs and the HRQL indexes that underlie them have been estimated for many illnesses, whereas credible willingness to pay estimates are available for relatively few illnesses. Using these estimates has the potential to approximate WTP when direct estimates are not available.

The first proxy measure would weight VSL by estimates of HRQL. To apply this approach, EPA should develop criteria for selecting (1) the health endpoints to value with the approach, and (2) the HRQL measures to apply. Criteria for selecting HRQL measures should consider the credibility of existing HRQL measures and their consistency with EPA's use of the VSL; existing "best practice" guidelines (IOM 2006, HHS 2016); methods of elicitation of HRQL (standard gamble, time tradeoff, rating scale, or other system); the sources providing the ratings (general population, patients, experts) relative to the population at risk for the illness considered; and how to choose a best estimate or to pool estimates using a suitable form of meta-analysis when there are multiple measures of HRQL for a single condition.

The second proxy measure would use monetized QALY values, that combine estimates of HRQL and duration, based on a value per statistical life year (VSLY) or value per QALY (vQALY) estimate derived from a VSL estimate. HHS and others use these estimates as rough proxies for WTP (see U.S. HHS 2016, U.S. HHS 2021).

Interpreting these two types of proxy measures as exact measures of WTP to avoid nonfatal illnesses is believed to require additional restrictions on the structure of preferences beyond those embedded in conventional welfare measures (e.g., Hammitt 2017a, Klose 2003). More research is needed to fully characterize these restrictions. Nevertheless, the SAB believes that proxy measures of WTP based on HRQL indices and QALY estimates may provide more accurate estimates of WTP than proxy measures based on COI alone, by addressing impacts on health-related quality of life that are not included in COI measures. Reporting benefit measures based on HRQL indices, QALY estimates, and COI estimates in a sensitivity analysis would usefully clarify the economic importance of choosing between these approximations.

The third approach the SAB suggests for EPA's consideration to improve the valuation of nonfatal health effects is structural benefit transfer (or "preference calibration," e.g., Smith et al. 2002, Smith et al. 2006) more generally. (Both of the approaches discussed above can be considered forms of such structural preference calibration.) This approach involves estimating a series of functions that relate an individual's WTP to characteristics such as their age and income. The functions are based on parametric utility functions and are calibrated to be consistent with available WTP estimates. As the functions are derived from an underlying utility function, the estimates are consistent with theoretical restrictions; for example, that estimates of WTP do not exceed individual income and are responsive to changes in the size of the risk reduction. Data drawn from several sources can be used to estimate

these functions, which can be applied to predict the variation in individual WTP attributable to changes in the modeled parameters. Newbold (2011) illustrates how a simple structural preference function can be used to compute WTP from COI and QALY estimates. Kenkel (2006) shows how the value of a statistical case of chronic illness can be derived by combining HRQL and VSL in a one-period model. However, his model implicitly assumes that the individual would be in full health (HRQL=1) without the condition, whereas many affected by air pollution-related morbidity are in less than full health, at least in part because baseline health declines with age [see, for example, Hanmer et al. (2006)]. The advantages of such models are that they can combine data from different studies that do not necessarily address the same outcome and ensure both internal consistency and consistency with theory. Thus, they are less reliant on completion of studies that use the same outcome measure and can be used to predict values in contexts outside of those studied. The disadvantage is that they require the analyst to assume an appropriate valuation function and calibrate its parameters. We recommend that EPA further investigate the use of these models and their potential application to value the impacts of air pollution on health.

In the long-term, EPA should encourage more primary WTP research for nonfatal illnesses. We recognize that stated preference (survey) research or revealed preference studies would be needed to develop these estimates. For example, Cameron and DeShazo (2013) demonstrate how stated preference research designs can be tailored to estimate WTP to reduce morbidity risks; numerous other examples are provided in the reviews listed above. Stated preference research conducted or funded by EPA requires OMB clearance under the Paperwork Reduction Act. Gaining such clearance is often very difficult. We encourage EPA to work with OMB to expedite the clearance of such research.

> **2.4.a.** *EPA currently uses a "20-year segmented lag" when discounting the value of PM2.5 and ozone-related deaths associated with long-term exposure, consistent with SAB advice received in 2004. EPA believes that a lag of five years or less as the primary method of analysis for PM2.5 and as a sensitivity analysis for ozone is more appropriate in light of the current best available scientific evidence. Please comment on whether EPA's conclusion reflects the current state of the science and whether an alternative lag structure associated with long-term exposure should be assumed.*

The SAB **recommends that the EPA continue to use the 20-year segmented lag rather than moving to a 5-year lag structure (Tier 1).**

The change from a 20-year segmented lag to a significantly shorter time frame (five years or less) of long-term exposure to either ozone or PM2.5 to model pollutant-attributable mortality does not seem based on a strong rationale attached to any new data or new analysis of previously collected data. Prior reports from the Advisory Council on Clean Air Compliance Analysis (e.g., Cameron 2001 and Hammitt 2008) suggest that the greatest mortality effects of PM reduction occur within the first 5 years of exposure, and the SAB concludes that evidence still supports this conclusion but that this is not inconsistent with longer term effects also being present. In the case of PM2.5, Section 6.4 of the TSD appears to cite Crouse et al. (2020) and Lepeule et al. (2012) as the basis for the EPA's current belief in the need for a shorter lag period. Crouse et al. (2020) analyze a national sample of Canadian adults and find that the hazard ratio describing the strength of the association between PM2.5 and non-accidental mortality increases monotonically as the duration of the exposure measure increases

from a lagged one-year average to a lagged three-year moving average to a lagged eight-year moving average. Lepeule et al. (2012) conduct an extended follow-up of the Harvard Six Cities study and find little difference in hazard ratios for lagged moving averages between one and five years. They do not consider periods longer than five years (see their Supplemental Material, Table 1). There are no new ozone data. It is unclear how the evidence from these two studies would rule out lag structures longer than five years. It is also unclear how and why the EPA believes that this evidence would be more supportive of a five-year lag structure than the EPA's current 20-year lag structure, which the SAB notes assumes that the majority of impacts (80%) takes place within the first 5 years. Restricting analyses to a 5-year lag also ignores latent effects including effects of early childhood exposure to PM that can affect health into adolescence and adult life. The change to a lag of 5 years or less would also seem inconsistent with new approaches EPA itself is developing to valuing benefits of PM2.5 reduction for non-fatal lung cancer, based on latency of 5-20 years between exposure and disease onset (Sections 5.3.6 and 6.4.2 in the TSD). In addition, in the response to charge question 4c, the use of latency from exposure to lung cancer onset may not be the same as the latency to onset of benefits. Latency between exposure and lung cancer onset may be long, but the decline in rates after exposure cessation may be much shorter (with persistent declines in risk during many years), based on evidence from cigarette smoking cessation studies. In addition, the latency from exposure reduction to achieved benefits may vary by chronic disease outcome. There is a relatively robust literature showing latency from chronic air pollution exposure to disease onset, but few studies have examined specifically the time course of benefits after exposure cessation. These are considerations for future research, and the SAB encourages the EPA to monitor the air pollution literature for studies that would allow greater certainty in assessing appropriate lags for benefits after exposure reduction.

Considering subgroups that may have exquisite susceptibilities to pollutants, e.g., the aged, those with cardiac or respiratory comorbidities, those with chronic inflammatory conditions or immunodeficiencies, etc., a compelling argument exists that the mortality risk in those populations occur in shorter timescales (Hill et al., 2023), but would be captured in the current segmented distribution. In summary, the SAB is unaware of compelling evidence supporting a lag restricted to five years or less when discounting the benefits of reduced PM2.5 and ozone-related deaths associated with long-term exposure.

> **2.4.b.** *Please comment on the approach to estimating neurological health effects using a cost of illness measure for hospital visits, as well as whether it is appropriate to assume that the first hospital visit indicates disease onset.*

The SAB discusses cross-cutting issues and recommendations related to the use of COI estimates at the beginning of the responses to Charge Question 4. In addition, the SAB concludes that the first hospital visit is unlikely to indicate the timing of disease onset, but that developing an improved measure of onset would require additional research. Regarding COI, as stated above, we **recommend that the EPA utilize more comprehensive estimates of averted costs beyond the cost of a patient's first hospital admission (Tier 1) while moving towards a WTP approach for valuing neurological and other health effects (Tier 1).**

*First Hospital Admission and Disease Onset:*

The TSD proposes to measure nervous system effects of PM2.5 based on first hospital admissions for Alzheimer's disease and Parkinson's disease. A hospital admission is a response to disease manifestation, and these diseases are not thought to occur suddenly, so the first admission is likely to occur subsequent to disease onset. This means it is not necessarily appropriate to assume that the first hospital admission indicates onset. The implications of this assumption for measuring the benefits of reducing PM2.5 depend on several factors not discussed in the TSD. These factors include the effects of PM2.5 on disease onset and progression, the shares of patients with each disease who are ever admitted to hospitals with those disease codes recorded as primary or secondary diagnoses, and the time between onset and first hospital admission. For reference, the Alzheimer's Association (2023, p.68) reports that 32% of Medicare beneficiaries with Alzheimer's or other dementias have at least one hospital discharge annually and that 22% of their hospital stays are followed by a readmission within 30 days.

The TSD constructs measures of neurological health effects from associations between PM2.5 exposures and first hospital admissions reported in Kioumourtzoglou et al. (2016). On page 27 of that article, the authors explain that "Given the design of our study and the use of administrative data, we were not able to assess whether air pollution was associated with the onset of neurodegeneration…our findings indicate that air pollution likely accelerates the progression of neurodegeneration, potentially after the onset of disease."

Evaluating how changes in PM2.5 exposure affect disease onset would require additional literature review or additional research. In principle, clinical diagnoses of Alzheimer's disease and Parkinson's disease could provide indicators of disease that would be more inclusive than hospital admissions and may be observed closer to disease onset. The U.S. Centers for Medicare and Medicaid Services (CMS) has criteria for translating medical claims into flags indicating disease presence. These data are available in the Center for Medicare & Medicaid Services' Chronic Condition Data Warehouse files (Center for Medicare & Medicaid Services).

*Cost-of-Illness:*
The COI measures for Alzheimer's and Parkinson's diseases presented in Section 5 of the TSD are based on hospital stays for a patient's first admission with the disease. These measures are likely to understate the benefits of reducing neurodegeneration for several reasons, as introduced in the discussion of cross-cutting issues above. First, patients experiencing neurodegeneration may have multiple hospital stays. As noted above, the Alzheimer's Association (2023) reports that among Medicare beneficiaries with Alzheimer's or other dementias, 22% of hospital stays are followed by a readmission within 30 days. Second, the progression of neurodegeneration is likely to generate additional medical costs before and after the first hospital stay, as well as costs for patients who are not admitted to a hospital. These additional costs include prescription drugs, home health care, nursing home stays, and other costs considered in TSD Section 6.4.5. Third, a potentially large cost not considered in the TSD is the shadow value of time associated with unpaid caregiver labor hours. For example, the Alzheimer's Association (2023) reports 11.5 million hours of time that unpaid caregivers, such as family and friends, devoted to caring for patients with Alzheimer's or other dementias in 2022. Fourth, neurodegeneration is thought to impair financial decision making, for example, by causing changes in portfolio allocation, changes in risk behavior, increased risk of financial mistakes, and increased vulnerability to financial fraud (Chandra et al., 2023).

Finally, COI measures, and estimates of averted costs more generally, include both out-of-pocket costs paid by the patient and costs paid by third parties such as insurers. As noted earlier, out-of-pocket costs likely understate what people would be willing to pay to reduce their risk of neurodegeneration given its nonpecuniary as well as its pecuniary effects. As recommended above, EPA should conduct a criteria-driven review of the WTP literature to determine whether they provide estimates that would be a reasonable supplement to the COI measures, either as part of the primary benefits estimates or in sensitivity analysis. In the long-term, we recommend that EPA encourage more research on WTP to reduce the risk, and slow the progression, of Alzheimer's and Parkinson's diseases as well as other health endpoints.

> **2.4.c**. *EPA currently values lung cancer using an estimated latency period between air pollution exposure and disease onset. Do you have any suggestions for improving the approach used?*

While we stressed the need to move beyond COI for all morbidity, here we specifically r**ecommend that EPA move beyond COI for lung cancer valuation (Tier 1) and examine whether the response to increases of air pollution exposure is equal but opposite to the response to equivalent exposure decreases for lung cancer occurrences (Tier 3).**

In order to value the reduction in mortality of reducing PM, Section 6.4.2 of the TSD states that EPA uses a 20-year distributed lag model approach to estimate the temporal distribution of reductions in mortality risk after reducing exposure, based on guidance from the Advisory Council on Clean Air Compliance Analysis in 2004 (Ostro and Cameron, 2004). Use of this lag was addressed in charge question 4a. Lung cancer has a latency between PM exposure and disease, and this has been used to develop models to estimate the temporal distribution of benefits after exposure reduction. To value the non-fatal lung cancer reduction benefits of reducing PM2.5, three models were developed by EPA with different latency distributions: An age-at-diagnosis cessation lag distribution model accounting for a lag-specific latency period; an adapted 20-year distributed lag model, using a 10-year modal latency (with a range of 5-20), based on 5 studies described in Section 5.3.6; and a latency-based triangular distribution of latency between 5 and 20 years. Each approach has some limitations. The age-of-diagnosis distribution, for example, assigns the highest reductions to the first 5 years of the distribution and not all ages have latency periods. The triangular distribution and adapted 20-year distributed-age distribution approach use the same latency period for all ages. There is merit in further investigation to identify age-specific latencies. A recent detailed cross platform ecologic study associating genetically-driven lung cancer incidence (EGFR-positive adenocarcinoma) in United Kingdom, South Korea, Taiwan and Canada with ambient air pollution measurements provides some guidance for specific classes of lung malignancies. Non-smokers living in high PM2.5 areas showed incidence coefficients ranging from .63 to 1.82. The relative incidence of these tumors (comparing high vs low pollutant exposures) was significantly higher at 3 years of cumulative exposure but not at 20 years. These findings suggest that for specific subsets of lung malignancies, plausibly those not exclusively driven by cigarette smoke exposure, a lag of 5 years or less is reasonable. However, given the broader spectrum of malignancies addressed in the charge question, these data do not support a change in lag/latency time (Hill et al., 2023). Nevertheless, these are reasonable approaches given the limited number of studies available to estimate latency to development of lung cancer and the lack of a good alternative.

An alternative and likely better approach would be to examine the reduction in risk of lung cancer after reduction of exposure, rather than a benefit estimated based on the latency to onset. The implicit assumption of the current EPA approach seems to be that in the absence of data showing the reduction in risk after reduced exposure that this will be equivalent in magnitude (opposite in sign) to the increase in risk with increased exposure. The reduction in risk of lung cancer after cessation of smoking has been established, and it is noteworthy that the distribution of reduction in risk is largest in the first 5-10 years after cessation, even though the latency to onset after exposure may be considerably longer in population studies, and in addition continues to decline during more than 30 years of follow-up. The temporal pattern of declining risk after smoking cessation is consistent in the American Cancer Study II cohort, worker populations, and in health conscious Seventh Day Adventists among whom smoking is unusual (and from whom the CR function for lung cancer used by EPA is obtained) (Markowitz et al., 2013; Gharibvand et al., 2017). Tobacco smoke is a much heavier PM concentration among smokers, and the composition is very different than ambient PM, but it would be worth evaluating whether the pattern of distribution of benefits from reduced risk after smoking cessation could be relevant to estimating the benefits of reduced PM exposure. How the empirical temporal distribution of benefits of smoking cessation compares with the latency to smoking-induced lung cancer onset could provide insight into the validity of the use of the current EPA approach to valuing lung cancer reduction. New research is needed to determine the distribution of lung cancer risk reduction after PM2.5 exposure reductions, including perhaps populations that move to areas with substantial differences in exposure.

The approach used to value non-fatal lung cancer in the TSD involves combining the latency period with an estimate of medical costs. We recommend that the Agency move away from the use of avoided illness costs as measures of benefits for nonfatal effects. Although most available WTP-based estimates of the value of avoiding a statistical case of cancer do not isolate the value of a nonfatal case, proxy measures based on either monetized QALYs or HRQL-weighted VSL suggest a value an order of magnitude larger than the approximately $33,000 proposed in the TSD. Guidance from the Nuclear Regulatory Commission (2023) uses monetized QALYs to value a case of non-specific, non-fatal cancer at $340,000 (2014) with a range of $130,000 to $630,000. (Nuclear Regulatory Commission, 2023). A meta-analysis by Sturza (2010) estimated HRQL weights for nonfatal non-small-cell lung cancer of 0.573 (metastatic) to 0.823 (non-metastatic).

An additional concern with the current EPA approach is that it assumes people retire at 65. The average retirement age is 65. However, it has been increasing over time, resulting in approximately 40 percent of the current workforce working past 65, so that valuation of lung cancer impacts should include lost wages for a more senior workforce as well as for the younger 'working-age' population.

> **2.4.d**. *EPA currently values acute myocardial infarctions (AMI) using a cost of illness measure. The cost of illness varies by age. Do you have any suggestions for improving the approach used?*

The SAB discusses cross-cutting issues and recommendations related to the use of COI estimates at the beginning of the responses to Charge Question 4 where we recommend that the EPA examine newer literature and develop a more current cost estimate including both averted medical and nonmedical costs. **With respect to AMI specifically**, **here we recommend that EPA should consider more recent health-labor studies that examine effects of AMI on labor market outcomes while relying on its value**

**of time guidance (U.S. EPA 2020a), which addresses the valuation of both market and nonmarket (labor and leisure) time (Tier 2).**

In the absence of applicable estimates of WTP to reduce risk of AMI, the EPA proposes to use avoided COI. Estimates of 3-year medical costs, exclusive of costs of initial hospitalization, are combined with estimates of reduced earnings over 5 years based on Cropper and Krupnick (1999). It would be better for the EPA to avoid reliance on studies this old, given that medical treatment costs and labor market conditions have changed significantly since that time. Older studies are also unlikely to reflect current standards for empirical best practices. The EPA should consider more recent health-labor studies that examine effects of AMI on labor market outcomes. For example, Stephens and Toohey (2022) find labor market effects 6 years after AMI, which is longer than the 5 years described in the TSD. As recommended earlier, EPA should also consistently rely on its value of time guidance (U.S. EPA 2020a), which addresses the valuation of both market and nonmarket (labor and leisure) time.

> **2.4.e.** *EPA uses morbidity cost of illness (COI) approaches for allergic rhinitis, asthma symptoms (albuterol use), cardiac arrest, school loss days, stroke, and work loss days. Do you have any suggestions for improving the approach used?*

The SAB discusses cross-cutting issues and recommendations related to the use of COI estimates at the beginning of the responses to Charge Question 4. In addition, the SAB **recommends that the EPA expand the valuation of lost school days to children younger than 13 (Tier 1) and consider adjusting the valuation for cardiac arrest and stroke to account for lost earnings (Tier 2)**. Additional discussion below focuses on issues specific to these endpoints.

*Allergic rhinitis:*
To estimate the avoided cost of illness for cases of allergic rhinitis in persons under age 18 years, EPA proposes (p. 88 of the TSD) to rely on Soni's (2008) estimate of mean annual expenditures from the Medical Expenditure Panel Survey (MEPS) (Agency for Healthcare Research and Quality). MEPS includes expenditures by all sources including patients and insurance companies. Soni (2008) reported that over half of expenditures are for prescription medications and most of the rest is for outpatient services; over-the-counter medications are not included. The proposed value is $600 per person with hay fever aged 0-17, based on Soni's estimated mean expenditure of $434 for this age group in 2005$ (inflation adjustment presumably used the medical care subindex of the Consumer Price Index (CPI)). The Soni (2008) study is a good choice for measurement of medical expenses of allergic rhinitis, and it matches the health impact function relationship between long-term exposure and prevalence.

It may be useful for EPA to investigate whether a related approach, based on allergen immunotherapy (allergy shots or tablets) would provide a valid measure for the population considered. The money and time costs of allergen immunotherapy, which are substantial, could be compared to the effectiveness of the therapy in reducing symptoms. Meta-analyses (Nurmatov et al., 2017, Calderon et al., 2007) generally find both subcutaneous and sublingual immunotherapy to be effective in reducing symptoms and medication use. Hardin et al. (2021) estimate costs and cost-effectiveness, and general information on the treatment is available from the American Academy of Allergy, Asthma & Immunology.[4]

[4] https://www.aaaai.org/tools-for-the-public/conditions-library/allergies

*Asthma symptoms (albuterol use):*
COI estimates typically include the costs of medical treatments to prevent or alleviate the condition, as noted earlier. Costs of albuterol use may be considered expenditures that defend against asthma symptoms. The method proposed in the TSD (p. 89) to measure costs of albuterol actuation would be a reasonable way to estimate a portion of defensive expenditures for asthma, particularly if coupled with a measure of the effectiveness of albuterol use in reducing symptoms.

*Cardiac arrest:*
The EPA proposes to measure avoided costs of illness from reduced nonfatal cardiac arrest based on estimates of medical costs derived from O'Sullivan et al. (2011). Over three years, medical costs excluding hospitalization and costs incurred during the first month following the event are about $35,000 to $36,000 depending on discounting (p. 90 of the TSD). Over three-fourths of the costs occur in the first year, suggesting that increasing the number of years considered might have only a modest effect on the total (although costs are larger in the third year than in the second).

The O'Sullivan study appears to provide a reliable estimate of medical expenses but failing to include lost earnings caused by cardiac arrest implies understatement of the full costs of illness. We are unaware of credible estimates of earnings losses caused by cardiac arrest in the U.S. A Canadian study (Garland et al. 2019) used tax records to estimate differences in employment and earnings for persons admitted to hospital for cardiac arrest relative to matched controls, and found a 12% reduction in employment, and a 14% reduction in earnings among those still employed, three years after cardiac arrest. These finding suggest that cumulative indirect costs of cardiac arrest over three years may be roughly similar in magnitude to estimated three-year medical expenses. If so, the total costs of illness may be underestimated by approximately one-half, and the EPA could consider adjusting costs to reflect this.

*School loss days:*
The EPA's proposed measure of the value of school loss days includes: (1) the reduction in lifetime earnings caused by the loss of learning due to school absences for middle- and high-school children only, plus (2) the value of caregiver time for all preschool and school children. The proposed approach to using results from Chetty et al. (2014) and Liu et al. (2021) to estimate the reduction in lifetime income caused by the loss of learning resulting from unscheduled school absences, as described on TSD pages 93-94, is broadly consistent with recent economic literature on education. However, it is unclear why the EPA proposes to restrict its measurement of lost learning to children ages 13-18. The literature shows similar effects of lost learning for younger elementary school students who take standardized exams during grades three through five. For example, Aucejo and Romano (2016) use a research design similar to Liu et al. (2021) and find that ten student absences reduce end-of-year math scores by 5.5%. Note also that Chetty et al. (2014) use data beginning in grade three, and the estimated association between ozone and absences from Gilliland et al. (2001) is based on 4$^{th}$-grade children, as described on TSD pages 55-56.

Against this background, the EPA should consider extending its current approach to valuing school loss days to include learning losses during elementary school. For example, the EPA could combine the results from Aucejo and Romano (2016) and Chetty et al. (2014) to develop lost learning measures for

elementary school students that would complement the existing measures that the EPA proposes to use for middle school and high school students.

The proposed estimate of the value of caregiver time for working parents is consistent with EPA guidance (US EPA 2020a), but the estimate for working parents omits the value of fringe benefits voluntarily paid by employers in apparent contradiction to the guidance. The reason for this discrepancy should be explained in the TSD or the value of fringe benefits should be included.

Two parts of the TSD description of the valuation of school loss days warrant clarification. First, the TSD (p. 93) lists four reasons why the proposed unit value is likely to understate the actual value of a school loss day. The first reason listed (omitting the value of avoiding the underlying symptoms or illness causing the absence) is valid. It is less clear what is meant by the second reason (omitting the opportunity cost of time for non-working caregivers) because the unit value estimate includes a measure of this value. We suggest the Agency remove this second reason or clarify what is missing from the unit value estimates. The third reason, omission of other aspects of school attendance such as social and emotional development or meals is valid. The fourth reason listed (not accounting for loss of learning in other subjects) may reflect that the Liu et al. (2021) estimate of the value of learning loss is based on missing classes in mathematics. As this aspect of the estimate is not stated in the TSD, we suggest the Agency clarify which 'other' subjects are referred to in their fourth reason. Second, Table 21 indicates the value is based on lost productivity of the parent, when in fact the larger share of the value is the impact of loss of learning on earnings.

*Stroke:*
The EPA proposes to measure avoided COI from reduced incidence of nonfatal stroke based on estimates of medical costs in the first year following the event derived from Mu et al. (2017). Estimates of longer-term costs could not be included because information on the timing of the costs was unavailable. The TSD (p. 94) cites evidence that about three-quarters of direct medical costs occur in the first year. The resulting cost estimate of about $34,000 is comparable to estimates of medical costs in days 31-365 following acute ischemic stroke from Johnson et al. (2016).

Avoided medical costs understate the full costs of illness for stroke by omitting indirect costs. We are unaware of credible estimates of earnings losses from incidence of stroke in the US. A Canadian study (Garland et al. 2019) used tax records to estimate differences in employment and earnings for persons admitted to hospital for stroke relative to matched controls, and found approximately a 19% reduction in employment, and a 19% reduction in earnings among those still employed, three years after stroke. These finding suggest that indirect costs of stroke may be roughly comparable to the estimated medical expenses. If so, the EPA estimate may understate the total costs of illness for stroke by about one-half and the Agency could consider adjusting costs to reflect this.

*Work loss days:*
According to the TSD (p. 94), work loss days are valued based on median annual earnings by county, exclusive of fringe benefits and employer overhead costs. In contrast, the value of lost working time for employed caregivers in the case of school loss days is based on national average earnings, inclusive of benefits and overhead. The TSD does not explain why the approach to valuing lost working time differs between the two apparently similar situations. Based on the values reported in the TSD, the different

approaches may cause a two-fold difference in the value of a lost workday. In the case of school loss days, the approach used to value caregiver time for working parents is consistent with the EPA guidance (US EPA 2020a). But the estimate for work loss days appears inconsistent with the guidance. If there are compelling reasons for differences in the way working time is valued in different situations, they should be explained; otherwise, a consistent approach to valuing working time should be utilized. In any case, if there is a considerable degree of monopsony power in the labor market (Council of Economic Advisers 2022), these measures would be expected to understate the value of lost production.

*Summary*
As discussed above, the averted costs associated with reducing the risk each of the endpoints/occurrences listed in question 4e is likely to be underestimated using the methods proposed in the TSD, for reasons including the omission of indirect costs of illness and the exclusion of the value of lost learning for children younger than age 13, as well as lack of estimates of willingness to pay (see discussion at beginning of responses to Question 4).

> **2.4.f.** *Do you have suggestions for improving the other morbidity willingness to pay (WTP) approaches: asthma symptoms (cough, wheeze, chest tightness, and shortness of breath) and minor restricted activity days (MRADs).*

The SAB has no specific suggestions for these endpoints but, as discussed at the beginning of question 4, we **recommend that the EPA undertake a criteria-driven review of the WTP literature (Tier 1).**

*Asthma symptoms (cough, wheeze, chest tightness, and shortness of breath)*
The EPA proposes to use the Dickie and Messman (2004) stated-preference survey to value reductions in symptoms for persons aged 0-17 years. While this study provides morbidity values for children's symptoms, two disadvantages are that the survey was conducted (1) in a single city and (2) in 2000. Concerning representativeness, Dickie and Messman (2004) provide evidence indicating that the study's estimates of parents' WTP to reduce their own symptoms were broadly similar to estimates from previous symptom valuation studies of adults. As discussed earlier, EPA should conduct a criteria-driven review of the WTP literature, including WTP studies for adults and children for similar health outcomes, which might allow additional data to be included. As part of this review, EPA may wish to consider using the ratios for nonfatal risks in Robinson et al. (2019b) to adjust adult values to estimate values for children in sensitivity analysis.

*Minor restricted (respiratory) activity days (MRRADs)*
There are no peer-reviewed studies providing an estimate of willingness to pay to avoid a MRRAD, and there might be any number of underlying respiratory symptoms that would cause an individual to restrict activities. The approach proposed in the TSD to determining a value for a MRRAD is reasonable in the absence of WTP estimates. However, MRRAD should be one of the endpoints considered in the WTP literature review referenced at the beginning of the responses to question 4.

> **2.4.g**. *Should EPA consider other morbidity or economic impacts of PM2.5 and ozone pollution, including but not limited to labor productivity and human capital formation? Which ones should be considered and under what conditions or in what settings they should be considered?*

The SAB **recommends that the EPA include the effects of air pollution exposure on both labor productivity and human capital using a WTP approach for valuation (Tier 1).**

It is important to include the effects on labor productivity and human capital. Changes in labor productivity and human capital from pollution likely reflect subclinical and, in some cases, transient health impacts. Given how difficult it is to measure and monetize these impacts, using these economic outcomes represents an important step for doing so. Furthermore, a now robust literature has emerged in the past decade to demonstrate consistent and significant effects from air pollution exposure on these outcomes (Aguilar-Gomez et al., 2022).

One question that arises is the extent to which these estimates may double-count impacts included in the values applied to individual health outcomes. As noted above, at times COI estimates include time losses due to illness. Whether WTP estimates include lost labor or leisure time depends in part on the study design, although conceptually WTP includes both pecuniary and non-pecuniary effects. Otherwise, the direct changes in productivity are in most cases an additional damage to include in the calculations and can be valued using prevailing wages as discussed in Graff-Zivin and Neidell, (2012) as well as the EPA (2020a) and HHS (2016, 2021) guidance on valuing changes in time use. Care needs to be taken to recognize the findings that demonstrate differential effects across sectors of the economy, i.e., outdoor workers are more affected than indoor ones.

Valuing the impacts on human capital can be more challenging, depending on whether the focus is on short-run changes in cognition or long-run changes in human capital. As long as care is taken to avoid double-counting, changes in human capital can be added to COI or WTP measures and monetized using evidence on the return to human capital (Currie et al., 2009). Short-run changes in cognition, however, are more challenging to value. Some cognitively demanding tasks can be substituted across days, for example, such that the welfare impacts from a cognitive insult may be minimal. Whereas some cognitive insults may have long-term effects if high stake decisions cannot be substituted intertemporally. One approach is to construct "minor cognitive-restricted activity days," which could be constructed to parallel the way the EPA uses results from the study by Chetty (2014) to translate test-score reductions (due to school absences) into reduced lifetime income. A focal study that credibly measures how PM2.5 affects test scores is Ebenstein et al. (2016), though there are some contextual differences from the U.S. At a minimum, the EPA can add a section to the TSD discussing cognitive effects on human capital production as a non-quantified benefit of reducing PM2.5 that can be updated in the future as the literature matures.

> ***2.4.h****. In some cases, EPA may determine that the evidence of a relationship between PM or ozone and a health endpoint indicates that the relationship is causal or likely causal but be unable to confidently value that endpoint (e.g., metabolic effects of short-term ozone exposures, for which the 2020 Ozone ISA made a "likely to be causal" determination). Can SAB recommend methods (e.g., bounding approaches) to better characterize presently unquantifiable health benefits for endpoints for which there is sufficient evidence of pollutant-attributable health effects?*

The SAB does not have recommendations for the inclusion of endpoints that cannot be confidently valued.

This question raised two issues: quantifying the health outcome in terms that are amenable to valuation and estimating values. The second issue, estimating values, is addressed at the beginning of the responses to question 4 as well as in the discussion of uncertainty under question 5. On the first issue, valuation is very difficult for outcomes that do not necessarily manifest in noticeable health effects. Individuals are unlikely to be able to provide valid responses to questions about WTP if they do not know how they are likely to be affected, nor can we estimate proxy values without this information. Ideally, these outcomes would be translated into the likelihood of occurrence for different health outcomes.

> **2.4.i.** *In some cases, EPA may determine that the evidence of a relationship between PM or ozone and a health endpoint is insufficient to indicate that the relationship is causal or likely causal. Section 1 describes a conceptual perspective on valuing health outcomes under causal uncertainty in the relationship between exposure and those outcomes. Are there scientific studies or practical settings where analysts accounted for causal uncertainty when estimating WTP estimates to reduce mortality or morbidity risks? Please comment on the appropriateness and potential for estimating and applying WTP estimates to avoided mortality or morbidity risks when there is causal uncertainty. An example is ozone-attributable all-cause mortality, which is discussed in section 6.2.3 of the TSD and not currently valued.*

The SAB **recommends that the EPA consider approaches for estimating WTP accounting for the full spectrum of uncertainty with respect to causality, including levels below the "likely causal" designation (Tier 1).**

On estimating WTP under causal uncertainty, there is an important conceptual distinction between expected WTP and WTP under uncertainty. The former would be computed by probability-weighting WTPs estimated under certainty; the latter would be computed as the compensating variation for some reduction in pollutant exposure conditional on the relevant uncertainties about the downstream health effects. Despite the important conceptual distinction between these two varieties of WTP, they should be close in magnitude when the risk changes being valued are small. To see this, note that causal uncertainty can be addressed in at least two ways, either in the main analysis or as part of a sensitivity analysis. First, a per-unit WTP measure estimated under the assumption of causal certainty can be scaled downward by an amount that reflects the uncertainty surrounding the causal influence of the pollutant on the health outcome in question. Such a scaling approach can be justified by the following logic. If the probability of causality is denoted as $p$, then the willingness-to-pay for a reduction in pollution from $x$ to $x - \Delta x$ by an individual with income $y$ and baseline health status $h$ would be defined implicitly by the following equivalence between expected utility without and with the pollution reduction,

$$\begin{aligned} U(y,h) &= \mathbb{E}\big[U\big(y - WTP, h(x - \Delta x)\big)\big] \\ &= pU\big(y - WTP, h(x - \Delta x)\big) + (1-p)U(y - WTP, h)\,, \end{aligned}$$

where $\mathbb{E}[\cdot]$ denotes the expected value. For small enough risk changes, a first-order approximation will accurately estimate expected utility, so we can re-write the second line of the expression above as

$$U(y,h) \approx p\big[U(y,h) - U_y WTP + U_h h_x \Delta x\big] + (1-p)\big[U(y,h) - U_y WTP\big],$$

which can be solved for willingness-to-pay under uncertainty:

$$\underbrace{WTP}_{\substack{\text{Willingness to pay} \\ \text{under uncertainty}}} \approx p \times \underbrace{\frac{\frac{\partial U}{\partial h}\frac{\partial h}{\partial x}}{\frac{\partial U}{\partial y}} \times \Delta x}_{\substack{\text{Willingness to pay} \\ \text{under certainty}}}$$

In words, for expected-utility maximizers facing small risk changes, WTP under uncertainty will be approximately equal to WTP estimated under certainty scaled by the probability of causality. Two important assumptions that must be maintained when using this approach are: 1) people's beliefs about causality and other aspects of uncertainty match the Agency's own characterizations of the scientific evidence about dose-response, and 2) people are not averse to ambiguity about causality.

This approach can be extended in a sensitivity analysis by scaling WTP by a range of values defined to reflect the uncertainty surrounding the causal influence of the pollutant on the health outcome in question. For example, the range attributed to a positive but causally-uncertain outcome could be defined to extend from a logical lower bound of zero (i.e., no true effect) to an upper bound defined by the high end of a 95% statistical confidence interval for a point estimate of the risk coefficient. This approach follows the general logic suggested by economic literature on partial identification, summarized in Manski (2008).

A second approach would be to use stated preference methods to elicit people's WTP under uncertainty by describing the state of the science to the survey respondents. The aim would be to elicit measures of WTP that embed the ways in which people value potential health improvements under causal uncertainty, attempting to estimate WTP under uncertainty on the left-hand side of the equation above directly rather than using the right-hand side approximation. This approach requires understanding peoples' beliefs. In the stated preference context, a review article by Johnston et al. (2017, pp. 328-329) on "Contemporary Guidance for Stated Preference Studies" recommends that practitioners communicate relevant information about causal uncertainty to survey respondents. In principle, the resulting WTP measures will reflect the importance that respondents attach to causal uncertainty. However, in practice, it may be challenging for analysts to effectively communicate causal uncertainty to survey respondents. Cameron et al. (2011) provide an example where survey respondents' answers to stated choice questions designed to elicit WTP to reduce morbidity and mortality risks appeared to embed the respondents' subjective adjustments to information they were given about the objective probabilities of experiencing particular health outcomes. Leggett (2002) outlines an analogous framework for adjusting revealed preference measures of WTP for divergences between objective information and the decision-makers' subjective beliefs.

Both approaches have advantages and disadvantages. The first approach may be more tractable and more broadly applicable given existing information available in extant non-market valuation studies. The second approach might be more accurate but would require conducting or commissioning one or more new tailored stated preference studies.

**2.4.j**. *Please comment on the income adjustment factors that EPA uses to value endpoints (section 5.4).*

The SAB **recommends that EPA update both the income adjustment factors and the base VSL to be consistent with the current literature within and across regulatory impact assessments. When estimating both VSL and values for nonfatal effects, the income adjustment should begin at the date when the data were collected, should rely on an income measure that is consistent with the income measures used to estimate the elasticities, and should use a formula that reflects currently accepted practices. For VSL, the Agency should apply an income elasticity of 1.0 with lower and upper bounds of 0.8 and 1.2, respectively in sensitivity analyses, and anchor the endpoints of the range at elasticities of 0.5 and 1.5 in probabilistic (Monte Carlo) modeling. For nonfatal effects, a central elasticity estimate of 1.0 should be used. (Tier 1)**

The SAB recommends updating the approach for income adjustments to reflect currently accepted practices, as described in more detail below. Applying these recommendations to the current VSL used by the EPA, which is based on 26 estimates from studies published before 1992 that draw from, on average, income data from 1978, would produce a central VSL estimate that is much higher than estimates drawn from more recent literature (e.g., U.S. Department of Transportation (DOT) 2021a, b; HHS 2016, 2021, Viscusi 2015, 2018). For example, updating the $8.7 million VSL (2015 dollars) reported on p. 84 of the TSD for inflation – based on the CPI-Urban deflator – yields a VSL of $10.8 million in 2022 dollars. Additionally adjusting for growth in real incomes using an income elasticity of 1.0 over 1978-2022 yields a VSL range of $21.8 to $22.7 million in 2022 dollars (depending on whether real GDP per capita or real disposable income is used as the income measure). To ensure that the VSL estimates used in EPA's analysis are reasonable given available empirical work, it seems essential that the EPA revise its base VSL rather than simply adjusting its current recommended default for changes in income.

The income adjustment factors presented in Table 32 and discussed in Section 5.4 of the TSD combine estimates of the change in real income over time (net of inflation) with estimates of income elasticity (i.e., the proportional change in WTP associated with a change in income). The discussion in section 5.4 is brief, referencing the BenMAP-CE User Manual (U.S. EPA 2018) for more detailed information. Because the 2018 version of the User Manual is not included in the TSD reference list and does not appear to be available online, we instead reviewed the 2022 edition and other EPA documents for more information on these adjustments.

The TSD provides income adjustment factors from 1990 through 2026, relying on estimates of per capita gross domestic product (GDP) to represent income. The TSD references suggest that the calculations reflect data available as of 2016. Monetary values are reported in 2015 dollars, adjusted for inflation using the GDP implicit price deflator (alternatively, at times the Consumer Price Index (CPI) is used in these adjustments). Historical GDP per capita estimates are from the U.S. Department of

Commerce's Bureau of Economic Analysis; future estimates combine GDP forecasts from the Congressional Budget Office[5] with Woods and Poole (2015) estimates of the projected change in the U.S. population. (Some limitations of the Woods and Poole data are discussed earlier in this report.) The approach in the TSD differs from the approach used by other major regulatory agencies (U.S. DOT 2021a,b; HHS 2016, 2021). These agencies rely on earnings measures rather than GDP to estimate income growth, and generally update their values annually to reflect the most recent data available.

Table 31 of the TSD presents a central estimate and lower and upper bounds for the income elasticities for minor health effects, severe and chronic health effects, and mortality. The TSD does not present a reference for the income elasticity values, although these appear to be values developed by Kleckner and Neumann (1999). As summarized in Ludwig and Neumann (2012) and Robinson and Hammitt (2015), the elasticity estimates for nonfatal effects reflect the range from eight studies published between 1975 and 1997; for mortality, the values are based on the range from seven studies and two literature reviews published between 1979 and 1998. These studies reflect data collected 25 or more years ago, which are not likely to reflect current conditions, data and methods now available, nor current standards for best practices. Several primary research studies as well as reviews and meta-analyses have been completed since the 1999 estimates were developed. DOT and HHS take these newer studies into account in estimating elasticities.

To combine estimates of changes in income and income elasticity to estimate values in each year, the TSD relies on a formula (p. 96) that relates proportional changes in WTP and income to the elasticity estimate. However, this formula is sensitive to the starting point. A more exact approach consistent with currently accepted practices involves using a logarithmic formula for calculating the arc elasticity over the income range considered.

*Time Period Covered by Adjustment*

For mortality, the EPA applies the income elasticity to its primary VSL measure. In the TSD, this is reported as $8.7 million in 2015 dollars (p. 84). It is difficult to compare this value to values reported elsewhere, because the estimates vary depending on the dollar year and the year in which the data on historical and future income were collected. However, as discussed in TSD Section 5.1.1, this estimate is consistent with the $4.8 million in 1990 dollars VSL used in its Section 812 reports on the retrospective and prospective benefits and costs of the Clean Air Act (U.S. EPA, 1997, 1999, 2011). That VSL is in turn derived from a review conducted by Viscusi (1992). The VSL estimate reflects an assessment of 26 estimates from 22 studies. While the level of the VSL is beyond the scope of this SAB's charge, a few characteristics of this set of studies have implications for the income elasticity adjustment. The set includes 21 estimates from labor market hedonic studies and 5 estimates from contingent valuation studies, and 5 of the 26 estimates address risk-income trade-offs in non-U.S. contexts. These studies were published between 1974 and 1991 and are based on data collected several years prior to their respective publication dates: specifically, income data collected in these studies date to 1978, on average.

[5] While the TSD states that future income estimates come from a Congressional Budget Office (CBO) 2016 publication, we could find no citation for CBO 2016 in the reference list. Appendix H-19 of the User's Manual states that future changes in annual income are based on data from the Annual Energy Outlook by the U.S. Energy Information Administration.

EPA's practice in applying this VSL in regulatory impact analysis has typically taken the following form: inflate the VSL to the common base year dollars used in the RIA and then adjust the VSL for growth in per capita income over a specified period of time. For example, the 2001 Diesel Sulfur Regulation (66 Federal Register 5002) RIA adjusts the VSL to 1999 base year dollars and then adjusts for changes in per capita income projected between 1990 and 2030 (EPA 2000, p. VII-33, VII-64), the focal year for presenting full implementation of the regulation. The 2014 Tier 3 Gasoline Regulation (79 Federal Register 23414) RIA adjusts the VSL to 2010 base year dollars and then adjusts for changes in per capita income over 2000-2030 (U.S. EPA, 2014, p. 8-16). Adjusting for growth in per capita incomes over different length time horizons – 1990-2030 in one case and 2000-2030 in another case – results in different measures of the VSL for the same point in time (year 2030).

Moreover, neither example accounts for adjustments that are merited to reflect the age of the publications. Ideally, the income adjustment should be applied from the date when the original data were collected in the studies underlying the agency's VSL to the target year in the regulatory impact analysis. In both examples above, the adjustment practice fails to account for changes in per capita income from the date of data collection to 1990 or 2000.

For nonfatal effects, the details of the approach for income adjustment are not discussed in the TSD. However, we assume the approach is analogous. As is the case for the VSL estimates, the approach should be consistent with the underlying studies. For example, if a study reports data collected in 1990, the income adjustment should be applied from 1990 to the target year. Note that income adjustments only apply to WTP estimates. WTP is sensitive to income because it reflects individuals' willingness to exchange their own income for changes in their own risks. COI estimates will also change over time, but the factors that affect medical and other averted costs differ from the factors that affect WTP estimates and hence must be addressed differently.

*Choice of Income Measure*

As discussed in Hammitt (2017b) and elsewhere, the choice of income measure can significantly affect the results of any VSL adjustment for income. The rate of change in labor earnings can differ significantly from the rate of change in GDP per capita, which is a much broader measure (see, for example, the comparison in Figure 4.1 of Robinson and Hammitt, 2015). One question is whether the income measure is consistent with how income was estimated in calculating the elasticity. For example, elasticity may vary depending on whether it was derived based on changes in earnings or in GDP per capita.

Another question is the consistency of income measures the EPA uses within and across regulatory impact analyses. For example, income projections are undertaken through the Shared Socioeconomic Pathways to produce future population estimates, as a part of the social cost of carbon (including adjustments of the value of reducing mortality risk associated with mitigating climate change risks), and in estimating the costs of proposed regulatory actions, as well as in updating the VSL for ozone and PM2.5 related mortality.

The EPA should clarify the measure of "income" used in different components of its analyses as well as the justification for selecting that measure.

*Choice of Income Elasticity*

Since the Kleckner and Neumann (1999) recommendations were developed, a significant number of primary research studies and reviews have been conducted that support a higher primary estimate for VSL income elasticity than 0.4. Examples of more recent reviews include reports developed for EPA (e.g., Ludwig and Neumann 2012; Robinson and Hammitt 2015) and for the World Bank (Narain and Sall 2016) as well as peer-reviewed journal articles (Hammitt and Robinson 2011, Robinson et al., 2019b) and several meta-analyses discussed below. In addition, both the U.S. DOT (2021a,b) and U.S. HHS (2016, 2021) use higher estimates.

More specifically, over 1997-2002, four studies reported income elasticities for the value of statistical life: 0.53 (Liu et al 1997), 0.89 (Miller 2000), 1.66 (Bowland and Beghin 2001), and 0.46 (Mrozek and Taylor 2002). Viscusi and Aldy (2003) replicated these studies and found that the vast majority of the variation in income elasticity reflected variations in the studies included in estimation samples. Their replications – based on a common sample of studies – yielded estimates across these four studies empirical strategies of 0.51 to 0.61, and further analyses in Viscusi and Aldy produce similar income elasticity estimates of 0.5 to 0.6. None of these five studies employed the same sample of studies as in the sample of 26 employed by EPA in estimating its VSL. Indeed, the entire VSL income elasticity literature draws on samples that differ from this set of 26. Some of this income elasticity literature draws on samples that include none of the EPA set of 26 studies because of reservations about systematic biases in measures of occupational fatality risk before the Bureau of Labor Statistics began producing the Census of Fatal Occupational Injuries in 1992 (see Viscusi 2004, Viscusi 2015).

More recent studies have employed methods for addressing publication selection bias in generating estimates of the VSL income elasticity. Doucouliagos et al (2014) conduct a broader meta-analysis of the VSL income elasticity literature, and in their evaluation of fourteen meta-analyses, they estimate a range for the income elasticity of 0.25 to 0.63. Viscusi (2015) conducts a meta-analysis of U.S. labor market hedonic studies that employ the Census of Fatal Occupational Injuries measure of occupational fatality risk and estimates an income elasticity of 0.8 to 1.1. Note that this measure of occupational fatality risk is considered the standard for labor market hedonic studies, but the Bureau of Labor Statistics began producing these data in 1992, after the most-recently published paper in EPA's sample of 26 studies used in estimating its VSL.

Viscusi and Masterman (2017) draw from a larger set of labor market hedonic studies – 68 studies across the U.S. and another thirteen countries – to estimate a U.S. income elasticity of 0.5 to 0.7, and a non-U.S. income elasticity of about 1.0. This is similar, and draws from a similar sample of studies, as the income elasticity estimated in Viscusi (2018) that varies from 0.53 to 0.85. In a review of the stated preference literature, Masterman and Viscusi (2018) estimate an income elasticity of 0.55 for "more affluent countries" (including the United States) and an elasticity of 1.0 for lower-income countries.

In addition to meta-analyses of VSL studies, Costa and Kahn (2004) estimated the value of statistical life through labor market hedonic analyses based on the decennial U.S. Census over 1940 to 1980. They estimate an income elasticity of 1.5 to 1.7. In a cross-sectional analysis of the U.S. labor market, Kniesner et al (2010) employ quantile regression and estimate how the compensating differential for occupational mortality risk varies with worker income. They estimate an income elasticity of 1.44.

In surveying this variation in income elasticities in the literature, DOT (2021b) adopted an income elasticity of 1.0. This decision reflected a weighing of the literature as of 2013, when the DOT estimates were developed. DOT references the 0.5 to 0.6 estimates in Viscusi and Aldy (2003), the 1.44 estimate in Kniesner et al.'s (2010) longitudinal study, and the 1.5 and 1.6 estimates in Costa and Kahn's (2004) cross-country analysis. HHS (2016) likewise adopted an income elasticity of 1.0, with similar references to Viscusi and Aldy (2003) on one end, and Kniesner et al (2010) and Viscusi (2015) on the other end. The International Monetary Fund has employed an income elasticity of 1.0 for its application of VSL to monetize the externalities from fossil fuel combustion (Parry et al. 2021), referencing Viscusi and Masterman (2017) and Robinson et al. (2019a). In its recent proposed updating of the social cost of greenhouse gases, EPA (2022c) has used an income elasticity of 1.0 for valuing mortality risk reductions, and likewise referenced, inter alia, Viscusi and Masterman (2017) and Robinson et al. (2019a).

We suggest that it would be appropriate for the EPA to use a central VSL income elasticity of 1.0. This reflects the balance of evidence across studies, recognizing that some produce estimates less than 1.0 and others produce estimate greater than 1.0. We note, with significant reservation, the inconsistency in applying a VSL income elasticity derived from the current literature to a VSL measure derived from a dated literature comprised of papers published before 1992. We also recognize that the income elasticity could enable income-based adjustments to valuing reductions in mortality risk across populations in a contemporary cross-section as well as across populations over time.

With respect to characterizing uncertainty in the income elasticity of the VSL (Tables 31 and 55 of the TSD), the SAB recommends adjusting the low and high estimates in light of the review of the literature with respect to charge question 4j. For the purpose of sensitivity analysis, 0.8 and 1.2 provide reasonable low and high estimates of VSL income elasticity. Note that these elasticity estimates should be applied to EPA's central VSL estimate. Combining two low- or two high-end values leads to extreme estimates that are highly unlikely; e.g., EPA should not combine its low elasticity and a low VSL estimate to develop a low-end value. For probabilistic (Monte Carlo) analysis, where a full distribution of values is needed, the ends of the ranges should be anchored at 0.5 and 1.5. An array of meta-analyses include 0.5 in their range (including Viscusi and Aldy 2003, Doucouliagos et al 2014, and Viscusi and Masterman 2017). Several studies estimated income elasticities in excess of 1.0, including the cross-sectional income elasticity of 1.44 from the U.S. labor market in Kniesner et al (2010).

For nonfatal risks, income elasticity has not been as well studied and may vary depending on the nature of the health effect. As discussed in Robinson and Hammitt (2015), the sparse evidence available suggests that there is substantial variation and uncertainty regarding the income elasticities appropriate for nonfatal health effects, due in part to issues related to study design. As a default, we recommend that the EPA apply the same estimates to nonfatal effects as to fatal effects and encourage further research on this topic.

*Combining Income and Elasticity Estimates*

The equation that the EPA uses to adjust for income changes (TSD p. 96) essentially assumes that income elasticity is a point estimate, derived by dividing the percent change in the VSL by the percent change in real income. However, this formula is sensitive to the starting point. A more exact approach involves using a logarithmic formula for calculating the arc elasticity over the income range considered.

Under this approach, the equation for adjusting VSL for changes in real income becomes the following, as described in more detail in HHS (2016, 2021) and elsewhere.

$$VSL_{(year\ y)} = VSL_{(year\ x)} * (real\ income_{(year\ y)} / real\ income_{(year\ x)})^{elasticity}$$

Where:

- $VSL_{(year\ y)}$ = the value of mortality risk reductions in the target year.
- $VSL_{(year\ x)}$ = the value of mortality risk reductions in the reference year.

We recommend that EPA change its approach to apply this formula.

## 2.5. Charge Question 5: Characterizing Uncertainty

*Charge Question 5: Section 6.4 and 6.5 describe approaches EPA uses for characterizing uncertainty. Please reference these sections in answering 5(a)-6(d).*

**2.5.a.** *EPA uses risk estimates from air pollution epidemiologic studies when quantifying the number of avoided cases of premature death and illness from future changes in air quality. Given that, by design, epidemiologic studies report associations between historical changes in air quality and the odds (or relative risk, or hazard ratio) of an outcome, how should EPA characterize this source of uncertainty? Are there methods to estimate the magnitude of this uncertainty that EPA should explore using? Please explain.*

The SAB is unaware of methods to reliably quantify uncertainties associated with the evolution of risk functions into the future. We **suggest that the EPA consider a retrospective analysis of the uncertainty associated with the historical changes in their own risk functions to provide a guide to the approximate magnitude of the uncertainty due to risk function changes over time (Tier 3).**

This question overlaps with other parts of question 5, for example, appropriate approaches to assessing the statistical uncertainty to the beta coefficient in the CR function, using standard errors to characterize sampling variation and heterogeneity of the estimates for non-sampling variation and simulation studies to link the CR function with benefits (question 5b). However, how health effects of historical changes in air quality reflect future effects (question 5a) is more challenging, as changes might occur that could affect the CR function. There may be changes in health and vulnerability of the population to pollution exposure. For example, there is already good evidence that the CR function is increased by co-exposure to COVID-19 (Marian et al., 2022; Chen et al., 2022; Chakraborty, et al., 2022) and this is not reflected in historical epidemiologic studies developing the CR function used in BenMAP. Mortality associated with PM2.5 may also be enhanced by climate change. There is uncertainty both to the change in temperature that will occur and to the CR function at different temperatures. Emerging evidence suggests that mortality may be greater during the co-occurrence of extreme temperature and short-term peaks of PM2.5 (Rayman et al., 2022), conditions that are becoming more frequent in areas where extreme temperatures also produce wildfires. Other anticipated changes that might affect the estimated benefits of regulation include both the level and the changing composition of pollutants as coal-fired power plants are retired and electrical vehicles become a larger proportion of the vehicular fleet.

In addition to uncertainty due to changing climate and sources of PM, there are recognized uncertainties that may have larger impacts on the estimated benefits. For example, as levels of PM2.5 fall to attain compliance with the current standards, the uncertainty of the shape of the mortality CR relationship at low levels will become more important for accurately assessing population benefits. As described in Section 6.5.15.1 of the TSD, the studies available are not inconsistent with a linear relationship at levels below 12 ug/m3, but there is some limited evidence from three Canadian studies, which provided key data at low levels for the most recent ISA, for both supra and threshold effects at lower levels (U.S. EPA, 2023). To complicate further the task of estimating benefits, deviations from a linear CR function could differ between PM2.5 and O3, and by outcome. In addition, relatively new methods for estimating population exposure to PM components and for assessing effects of pollutant mixtures may challenge the historical assumption that EPA and other regulatory agencies have made that PM2.5 toxicity depends exclusively on concentration with little variation by composition. Emerging results using rapidly evolving methods for assessing impact of pollutant mixtures will need to be incorporated into future ISAs. Advances in medical technology and treatment could also affect the CR function.

Existing studies on COVID-19 and PM2.5 interactions on mortality may be sufficient to estimate new population CR functions for PM2.5 and incorporate them into the ISA and BENMAP, though there is clearly uncertainty as to the duration of COVID-19-related effects on CR functions. Emerging research on PM composition and mixtures, applied to historical cohorts that form the foundation from which EPA estimates benefits, is likely to provide results useful for the revision of ISAs in the near future. However, additional research is likely needed to assess how historical CR functions may be affected by changing conditions like climate change for which there is a paucity of historical data to be able to estimate impacts on CR functions, or of electric vehicle penetration, for which health impacts are only beginning to be assessed (Garcia et al., 2023).

It could be informative to frame the uncertainty in projections of changes in CR functions over time by reporting the quantitative importance of the historical evolution of the EPA's preferred CR function. For example, the EPA could prepare a table summarizing the sensitivity of benefit estimates to a range of values for the CR hazard ratio, where that range is defined by the set of hazard ratios that the EPA has used in past regulatory analyses. For example, in the EPA's *Regulatory Impact Analysis for the Proposed Reconsideration of the National Ambient Air Quality Standards for Particulate Matter*[6] (U.S. EPA 2022a) the benefit measures in Table 8-1 of that document could be recalculated using the baseline hazard ratios that the EPA previously used when developing the *Second Prospective Study of the Clean Air Act Benefits and Costs*[7] (2011) and the First Prospective Study[8] (U.S. EPA 1999). Reporting the sensitivity of benefits and net benefits to changes in the EPA's preferred hazard ratio from 1999 to 2011 to 2023 could help readers to assess the potential magnitude of uncertainty under the assumption that future changes in hazard ratios are similar in magnitude to past changes. We note that such an assumption is speculative as the change in preferred CR functions over time may reflect factors such as changes in the health of the population or in the composition of particulate matter that would reflect genuine variation in the CR function but they might also reflect advances in data, pollution

[6] https://www.epa.gov/system/files/documents/2023-01/naaqs-pm_ria_proposed_2022-12.pdf

[7] https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2020-report-documents-and-graphics

[8] https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2010-first-prospective-study

monitoring, and/or statistical modeling techniques that have led to improved quantification of the CR function (which might or might not have varied in time). Even changes in factors that affect the CR function, such as the health of the population or the composition of particulate matter mentioned previously, might not continue to change with similar magnitudes in the future. Nevertheless, such a retrospective analysis might provide useful information about the scale of future CR function changes, which are difficult to forecast.

> **2.5.b.** *EPA performs a Monte Carlo simulation to quantify a distribution around certain input parameters including the beta coefficient used to quantify attributable cases of air pollution-related effects. Are there other input parameters for which EPA should specify a distribution? To the extent that insufficient data exist to specify such a distribution, should EPA assume a distribution? What other methods are available to quantitatively or qualitatively assess uncertainty if sufficient data are lacking?*

The SAB **recommends that the EPA improve the representation of uncertainty in their analysis by moving towards a more holistic treatment within the Monte Carlo simulation. Furthermore, the discussion of uncertainty in the TSD can and should be clarified (Tier 1).**

Section 6 of the TSD outlines a three-tiered strategy for dealing with uncertainty in their analysis: 1) Monte Carlo assessment of benefits associated with PM2.5 and ozone reductions (Sections 6.1 and 6.2); 2) quantitative sensitivity analysis on selected inputs (Section 6.4); and 3) qualitative discussion of remaining uncertainties that are difficult to quantify (Section 6.5). Section 6.3 describes a set of quality assurance steps on the baseline incidence data rather than a proper sensitivity or uncertainty analysis and, therefore, does not belong in Section 6 and should be moved to section 3 (probably section 3.2 as it applies to 3.2 and 3.3). This tiered strategy is fundamentally a sensible one given the complexity of the problem and the myriad sources of uncertainty, not all of which are well quantified.

However, the EPA should adopt a rigorous and holistic approach to evaluating uncertainty in its benefits assessments. From Section 6 of the TSD, it appears that the EPA currently considers a large number of sources of uncertainty but does so in a piecemeal, one-at-a-time manner rather than assessing an overall uncertainty from all sources combined. Monte Carlo methods are probably the most flexible and rigorous approach for doing this, but the EPA's analyses should support a Monte Carlo calculation that produces a single output distribution of benefits that combines sources of uncertainty across multiple inputs. We also encourage the Agency to explore the adequacy of existing data to evaluate the co-variance of different sources of uncertainty, which could then be included in the Monte Carlo analysis.

The three-tiered strategy currently used by the EPA still makes sense even if a holistic Monte Carlo approach is adopted. There will be sources of uncertainty that turn out to be small in the overall assessment. For these, a quantitative sensitivity analysis (Tier 2) can be performed without needing to incorporate them into a full Monte Carlo assessment. There will be other sources of uncertainty that cannot be quantified but should still be recognized and discussed (Tier 3).

For uncertainties where the EPA performs a Monte Carlo assessment (e.g., PM2.5 mortality and other endpoints examined in Sections 6.1 and 6.2), the EPA uses the reported standard errors from a given

epidemiological study to quantify uncertainty in the inputs. Standard errors account for sampling variation only and give an incomplete picture of overall uncertainty. A better approach would incorporate and account for heterogeneity of risk estimates across the entire literature. Meta-analyses, formal expert elicitation or other forms of expert judgement are methods to assess appropriate uncertainty distributions on key input parameters. Where limited data are available to assess uncertainty, simple distributions for key input parameters (e.g., uniform or triangular distributions) can be used and are preferable to ignoring these uncertainties altogether. Manski's work (2008) on partial identification on bounding assumptions also might be useful in this context.

The SAB identified several sources of uncertainty that ought to be considered. There is uncertainty in the air quality modeling that links emissions changes to changes in concentrations and exposure. This uncertainty is likely somewhat smaller than that for risk estimates and WTP but still appreciable. Literature suggests that exposure misclassification tends to bias reported hazard ratios downward leading to a low bias in risk estimates. Other limitations of epidemiological studies (e.g., whether it accounts for co-pollutants, temperature, and other stressors) contribute uncertainties to risk estimates. Uncertainty in assumed income elasticity on WTP is important in assessing benefits, and the range of elasticities currently used in the EPA's analyses is probably too low. Additional uncertainty accrues when the economics literature infers WTP from working-age adults, but these are necessarily applied to older populations who account for a large share of the quantified premature mortalities due to air pollution. Not all these sources of uncertainty are equally important or even possible to quantify, but many could be meaningfully incorporated into Monte Carlo simulations to give a more holistic assessment of uncertainty. Uncertainties in income elasticity of WTP and in the air quality modeling – and perhaps others – are probably tractable and important enough to specify a distribution using some form of expert judgement.

The TSD should be revised to clarify how key uncertainties are treated. The introduction to Section 6 is quite short and gives only a very limited overview of how uncertainty is handled. Much of the key information is found in Sections 6.1 to 6.5, which are lengthy. Essential information often gets lost amongst many less important details. It is difficult to determine what distributions are being used in the Monte Carlo analysis for key parameters such as the beta parameters and WTP/VSL. The introduction to Section 6 should be expanded to give a more detailed overview. This could include a more thorough description and justification of the overall approach for characterizing uncertainty, a list or table of which parameters get which tier of treatment (Monte Carlo versus quantitative sensitivity versus qualitative), and a summary of distributions for parameters that dominate the overall uncertainty in benefits.

> **2.5.c.** *Does this section sufficiently characterize the various sources of uncertainty associated with the estimate of WTP for reduced PM2.5 and ozone exposure? Are there other sources of uncertainty that warrant identification and evaluation?*

We conclude that this section appropriately characterizes the primary sources of uncertainty and **suggests several additional factors that warrant additional consideration (Tier 3)**.

*Income elasticity bounds:*

TSD Section 6.4.3 outlines sensitivity bounds on income elasticities. These bounds should be revised, as needed, to be consistent with any changes that are made in response to the SAB's comments on central estimates and ranges for income elasticities in the response to charge question 4j.

*Errors in measuring pollution exposures*
TSD Section 6.5.3 addresses measurement errors in assigning pollution exposures. While techniques for predicting pollution exposure have improved over time, significant challenges remain due, for example, to variation in building sealing and individual mobility through commuting and other activities. Studies that use techniques designed to avoid biases from measurement errors often find that addressing measurement errors (together with other sources of confounding) increases estimates for the causal effects of pollution exposures on health outcomes (Deryugina et al., 2019; Graff-Zivin and Neidell 2013; Aguilar-Gomez et al., 2022).

*Latency periods*
TSD Section 6.5.6 addresses the delay between changes in exposure and changes in health. Most latency effects last just a few years, though some persist longer, but even short delays could be critically important for childhood exposures.

*Confounding by individual risk factors*
TSD Section 6.5.9 addresses confounding by individual risk factors in epidemiological studies and states that it is not currently possible to quantify uncertainty due to this confounding. However, research designs have been developed to mitigate this source of confounding and these designs could be applied to epidemiological studies. Graff-Zivin and Neidell (2013) and Aguilar-Gomez et al. (2022) provide recent reviews of this literature.

> **2.5.d.** *Please comment on whether and how EPA could account for uncertainties associated with the shape of the concentration-response function, particularly at concentrations near and below the lowest concentrations observed in the epidemiologic studies used to quantify health benefits.*

We **recommend that the EPA incorporate results from additional CR functions beyond those that meet the criteria used in the ISAs to form a more robust pooled uncertainty estimate (Tier 1) and extend its current sensitivity analyses to incorporate results from studies of the association between mortality and short-term air pollution exposures (Tier 2)**.

First, the current approach to calculating and propagating uncertainties is not well explained in the documentation. It appears the uncertainties for each CR function are developed from the uncertainties garnered from the individual studies and then Monte Carlo simulations are used to provide the given uncertainties. It would seem this would be done analytically, assuming a given uncertainty distribution (which should be documented). Further, how they are propagated is not well explained, and a sample calculation should be shown, e.g., in an appendix or similar.

As noted elsewhere in this review, the EPA should include additional CR functions beyond those that strictly meet the criteria used in the ISAs, particularly the Global Exposure Mortality Model (Burnett et al., 2018)) should be supported. This CR function, based on data both from North America and data

from elsewhere, has a different shape. Including such functions would create a larger set of responses to be included in the overall pooled uncertainty estimate recommended previously by the SAB, providing additional information to more robustly characterize uncertainties, including at lower concentrations.

It would be instructive to have a box showing the CR functions supported by BenMAP (and potentially others), along with their uncertainties, in both the TSD and the BenMAP manual for the most important endpoints (e.g., mortality for ozone and PM, asthma incidence for ozone, etc.), graphically comparing the functions. The figure caption should note the limitations of such comparisons. For studies where different CR function shapes have been evaluated, a separate graphical presentation can be provided, focusing on the lower concentration range. While it is recognized that the EPA has chosen not to develop a single composite CR function, a comprehensive discussion of what the comparison of CR functions and their uncertainties suggests about the range of potential values over the likely range of concentration exposures would be instructive to individuals using the TSD and BenMAP. However, the EPA need not go into great detail given the limited, very specific applications of the TSD. Nevertheless, given the likely broader range of use for BenMAP, and to the extent that the TSD also provides documentation for BenMAP, some discussion is warranted. Potentially more detail could be contained in the BenMAP *User's Manual* and its Appendices.

The EPA could account for some of the uncertainty associated with the shape of the CR function at concentrations near and below the lowest concentrations observed in the epidemiologic studies by extending its current sensitivity analyses to incorporate results from studies of the association between mortality and short-term air pollution exposures. In principle, studies of short-term air pollution exposures could be used to identify a lower bound on the value of reducing air pollution in locations where the CR function for long-term exposure cannot be directly estimated. Such locations are still likely to experience short-term exposures that are within the range of concentrations observed in epidemiologic studies that estimate CR functions for short-term exposure. Valuing reductions in short-term exposure only in these locations would give a lower bound on the total value of reducing exposure under the EPA's maintained assumption that the association between long-term exposure and mortality is inclusive of (but not limited to) the effects of short-term exposure on health outcomes (TSD Table 1). In the economics literature, Deryugina et al. (2019) provide evidence supporting a causal relationship between short-term PM2.5 exposures and mortality among people over age 65. The upper uncertainty bound would remain that based on the pooled estimates from long-term exposure studies.

> **2.5.e.** *Please comment on any other assumptions that significantly impact the magnitude of the benefits for which EPA should generate more explicit assessments of uncertainty, either quantitative or qualitative.*

Regardless of whether or not the EPA updates the VSL to be more consistent with current literature, we **recommend that the EPA extend its current approach to quantifying sources of uncertainty in benefit measures to include a more explicit assessment of uncertainty associated with VSL including the shape of the VSL's uncertainty distribution and the transferability of VSL measures from EPA's *Guidelines* values to the population affected by changes in air pollution (Tier 1).**

One source of uncertainty arises from the distribution of VSL measures used. The 26 studies currently used by the EPA to construct a VSL distribution were published between 1974 and 1991. Since then, there have been many developments in data availability and econometric techniques. In Section 2.4.j, the SAB recommends updating the distribution of VSL studies used to reflect current literature. Regardless of whether or not the EPA updates the VSL distribution, sensitivity analysis using the ranges of estimates recommended by other major Federal regulatory agencies (DOT 2021a, b, HHS 2016, 2021) could be informative. A second source of uncertainty arises from differences between the populations that are used to estimate VSL in academic studies and the populations believed to be most affected by changes in air pollution. Most of the VSL studies cited in the EPA's *Guidelines for Preparing Economic Analyses* (U.S. EPA, 2014) are based on employment decisions made by workers aged 18 to 65. In contrast, most of the premature mortality from air pollution is concentrated among people over age 65. Thus, the VSL study population differs from the transfer population in terms of age, health, wealth, and other characteristics that could affect the population average willingness to pay to reduce mortality risk.

## 2.6. Charge Question 6: Data

*Questions 6-10 are related to the second goal of the panel (how the program implements EPA's methods to quantify estimated benefits, focusing on the user interface and software engineering), in the context of the recently developed cloud-based implementation of BenMAP. BenMAP is a complex tool incorporating air quality modeling results, demographic and population data and projections, economic valuation estimates, health related concentration response functions, and numerical methods and parameterization. To support the Panel's review of BenMAP, EPA is providing a document entitled "User's Manual"* (U.S. EPA , Dec. 2022)[9] *as well as providing panel members access to the cloud-based version of the tool and training on how to use the tool.*

*Charge Question 6: BenMAP uses a broad set of air quality (monitoring and modeling), demographic, economic, concentration-response, incidence, and prevalence data.*

***2.6.a.*** *Are the contents and sources of input data transparent and appropriate?*

The TSD and the BenMAP *User's Manual* have provided extensive documentation on the literature and preparation process for the default data choices in the online BenMAP, including air quality surfaces, incidence/prevalence rates, population datasets, health impact functions, socio-economic and demographic data, inflation, valuation functions, and income growth adjustments. In general, the contents and sources of data are described transparently. The appropriateness of the input data for use in BenMAP is harder to evaluate. Furthermore, the online tool doesn't provide direct access to some of the data used, such as demographic and baseline incidences, making the evaluation and user interpretation/diagnosis of the results difficult. To make it fully transparent, **we recommend that EPA include access to all the data elements that it has permission to distribute (Tier 1). We also recommend that users be able to upload their own datasets (Tier 1)**. These could also be provided for general use in the data center so that people can further diagnose their model outputs.

*AQ input choices:*

[9] https://sab.epa.gov/ords/sab/f?p=114:0:32679603711722:APPLICATION_PROCESS=AGENCY_REVIEW:::AR_ID:2489

The model offers a variety of pre-policy and post-policy air quality (AQ) surfaces data options, but it's not clear what each option is. Some explanation of what these are would be helpful. If these options are only for demonstration purposes, the process for the user to upload user-specific AQ data needs to be better documented. For example, the shape file of current 12 km grid definition needs to be provided for the user to convert a user defined air quality surface or other input data into the BenMAP grid definition. Given that the population data and baseline incidence rates are locked into the default spatial resolutions, such conversion of AQ surfaces is a necessary step. However, if future cloud versions of BenMAP can accommodate a flexible user defined population, the air quality surface will need to be prepared differently. Finally, the functionality to use monitoring data is not yet currently implemented in the online BenMAP. Following our recommendation above that future versions of the online BenMAP incorporate the functionality for users to use their own data, monitoring data could then be uploaded**.** Future versions may also consider including emerging spatial and temporal air quality surfaces, such as satellite data, or a low-cost sensor network such as PurpleAir.

*Population data and baseline incidence data:*
Open-source data are preferred. Until such time as open-source can be implemented, **we recommend evaluation of Woods & Poole data against the American Community Survey data from the U.S. Census[10] to benchmark its validity (Tier 2).** Finer grid resolution for population and baseline incidence rates are more appropriate for environmental justice applications. For example, population data at 100 m resolution is available in the literature (e.g. Depsky et al., 2022). Currently, users cannot (yet) enter their own population data or incidence data, only the pre-loaded U.S. population. This is especially important for incidence data because users may need to use their own validated local and regional health data that are more appropriate to their study domain. We also encourage the Agency to take advantage of related EPA work that has built tools to analyze air pollution sources and impacts at high spatial resolution, including environmental justice indicators, by working collaboratively with the NEXUS project (https://nexusweb-test.app.cloud.gov).

*Health effects studies:*
In the TSD, studies that fell outside of the inclusion criteria in Table 42 could be provided, for example, as external links, for users to evaluate the appropriateness to include or exclude. Since the papers are taken from the EPA's *Integrated Science Assessments*, which are quite robust, the evaluations are appropriate. For inputs into running BenMAP, one key limitation is in the choice of health effects studies. The user cannot select individual health effects, only the sets of all mortality or morbidity effects. Moreover, unlike in the desktop version, users cannot apply their own functions that are more appropriate for their study region, which limits the opportunities to customize the analysis. **We recommend that future versions of BenMAP allow users to apply their own health impact functions (Tier 1).**

As a cautionary note, although the criteria are clear for choosing health endpoints in terms of causality and biological plausibility and core preferred health effect studies, one challenge that arises is when these endpoints and preferred studies change over time – e.g., from the 2009 PM2.5 *Integrated Science Assessment* (U.S. EPA, 2009) to the 2019 *Integrated Science Assessment* (U.S. EPA, 2019). Health effects evaluated at different points in time will be confounded by changes in the selection

[10] https://www.census.gov/programs-surveys/acs

criteria. For example, when a user tries to evaluate air pollution attributable health impacts over time by comparisons to previous BenMAP studies, such a comparison becomes problematic as the previous studies are not compatible with the current version.

> ***2.6.b.*** *Are the inputs appropriately indicated for user selection and implemented as documented in the referenced literature?*

The documentation and explanation of the inputs in both the *User's Manual* and TSD are generally adequate. However, the online BenMAP could make it more intuitive for users to relate the options to their descriptions in the TSD. It was unfortunate that there wasn't information within the program itself to more deeply interrogate the inputs. **We recommend that the Agency add additional ways within the online BenMAP for users to interact with and understand the inputs and reference literature (Tier 2)**. Also, as mentioned above, the lack of input choices, particularly user-defined inputs, is a limitation of the online BenMAP at this point.

> ***2.6.c****. Is the information provided to judge the sources and quality of the underlying data used for calculation or as input to algorithms and parameterizations sufficient?*

The online BenMAP locks users into predefined analysis paths based on standardized exposure fields and high-quality published studies for CR functions. This "design feature" forces BenMAP users to use standard model results and published health response data for calculations, but more advanced researchers may need flexibility to explore a broader array of inputs and configurations. **We recommend that EPA provide explicit instructions and examples that describe how to introduce new input data and CR functions (Tier 2).** Examples and details follow below.

Selecting among the published CR functions is challenging for new and intermediate users because the subtle features of each study are difficult to differentiate. It would be helpful for EPA to publish their "expert rating" for each study based on the criteria listed in the offline documentation. EPA's comments and ratings in each category and an overall rating would greatly help users understand which CR functions may be most applicable for their own work. Alternatively, should the EPA support an expert elicitation on the credibility of various CR studies (see response to charge question 4) that information could be provided in BenMAP. It may also be helpful to develop a feedback feature where users can post comments and advice that can be accessed by new users. The combination of these features would guide new users towards the most appropriate datasets while still allowing the expert users the flexibility to test alternative datasets.

As a case study to illustrate the need to better explain how to judge the strengths of each CR function, it appears that the BenMAP model and the CO-Benefits Risk Assessment[11] (COBRA) model use different default CR functions for PM mortality. It isn't clear why these two EPA models made different choices. The rationale for the choice between CR functions should be easily understandable to all BenMAP users in the documentation.

[11] The CO-Benefits Risk Assessment Health Impacts Screening and Mapping Tool (COBRA) is described on EPA's website at https://www.epa.gov/cobra.

The procedure for a BenMAP user to replace model inputs with their own data needs more documentation with examples. Most importantly, this includes the ability for users to supply their own exposure fields developed through chemical transport model (CTM) studies and/or through the processing of monitoring data. Other inputs such as socio-economic data, demographic data, inflation rates, etc should also be placed in a data center with links to the online version of BenMAP. Performance metrics for future demographic predictions should be developed to help understand the accuracy of this data, especially in cases where proprietary datasets are used. Open-source data would be preferred for all inputs to BenMAP whenever feasible.

Providing documentation on the BenMAP source code would potentially create a new type of user that could fully exercise the capabilities of BenMAP and possibly even contribute new code to future versions of BenMAP; **thus we recommend that EPA work towards making BenMAP "open-source" in such a manner that researchers could contribute to source code in a managed collaborative process (Tier 1).** There are several "open-source" model projects that follow this paradigm, including the Community Multi-Scale Air Quality (CMAQ) and the Weather Research and Forecasting Model with Chemistry (WRF/Chem). Harnessing the creativity of the research community to grow BenMAP has the potential to greatly enhance the quality and functionality of the model.

> ***2.6.d**. Are the concentration-response functions appropriately implemented, documented, and referenced for users to evaluate and select? Are the concentration-response functions appropriately included or indicated in outputs?*

As mentioned previously, we would like to see BenMAP upgraded so that users can add their own health impact functions rather than being restricted to those preloaded into the model. For the choice of health effects studies and CR functions, the user cannot select individual health effects, only the sets of all mortality or morbidity effects. This limits the questions that researchers can answer, e.g., BenMAP users often want to examine how air pollution attributable health impacts change over time – in comparison to previous BenMAP studies – but this becomes problematic when the sets of morbidity effects change. For example, the 2009 and 2019 PM2.5 ISAs (U.S. EPA 2009, 2019) understandably made different determinations about biological plausibility and causality of health endpoints, and therefore BenMAP should allow users to select endpoints.

Where BenMAP provides a variety of CR functions for the same health endpoint, it should provide users with some description of the conditions under which that particular CR function was evaluated to aid users in identifying CR functions that are most appropriate for particular applications.

Another issue is that there is some inconsistency with other EPA models that examine health and economic impacts of PM2.5 pollution, e.g., EPA's COBRA model, a reduced form model that estimates health and economic impacts directly from criteria pollutant emissions changes. The latest version of COBRA uses a different set of health effect studies; notably it still uses for PM2.5 mortality results from the American Cancer Society and Harvard 6-Cities cohort studies. Perhaps this is a COBRA issue rather than a BenMAP issue. Still, this inconsistency across models is problematic and more flexibility in modeling choices would allow users to perform more consistent analyses.

For ozone, care needs to be taken that the air quality data input to BenMAP is consistent with the CR function used. For example, the daily maximum 8-hour ozone levels are provided but health effects are computed for 1-hour effects. There was limited documentation about how BenMAP converts between 1-hour max to 8-hour max and vice versa. This could be more transparent. It appears the conversion works both ways without loss of information, i.e., if you convert 8-hour to 1-hour and back, you end up with the original value. Better documentation would help the user navigate these issues.

Also, in terms of health endpoints for ozone, it was not clear how to adjust the CR functions for annual versus warm season ozone when quantifying acute health effects. For mortality, the studies are almost all from the warm season. For morbidity, some studies cover other seasons or the annual basis. These issues with averaging time for input ozone data need to be clear to the user and, ideally, BenMAP would be able to enforce consistency between the air quality data and CR function in terms of time period.

***2.6.e**. Has EPA appropriately characterized the capabilities and limitations of the BenMAP tool and supporting appendices with respect to incorporating externally created air quality modeling data inputs (3.3.1) and estimated changes in air quality over time? Are there tools that EPA might consider offering to improve information quality for air quality monitoring data inputs?*
*For reference, the document "User's Manual" describes the inputs [Chapters 2,3,5], underlying data [Chapter 11, Appendix I], concentration-response functions [Appendices B, D, E, F], and economic valuation parameters [Appendices G, H].*

We agreed that the standard capabilities for incorporating externally created air quality modeling data inputs (3.3.1) and estimated changes in air quality over time were adequately described by the online BenMAP tool and supporting appendices, including many planned capabilities not yet implemented. We identified documented limitations in using pre-loaded air quality data, uploading and formatting user-created data on the 12km grids, using monitoring data (not yet supported) as well as inconsistent and missing variable documentation and incomplete guidance on using the surfaces provided in the current version.

**We thus recommend improvements to the guidance provided for preparing air quality surfaces and using the surfaces currently available (Tier 1).** Algorithms for calculating each metric (e.g. "QuarterlyMean," "WarmSeason_D8HourMax") were not presented or provided via citations in the tool, appendices, or TSD document. Full documentation and functionality (in BenMAP or external scripts) to calculate BenMAP-ready surfaces for all metrics from an annual hourly BenMAP-formatted text file or CMAQ I/O API file would be helpful additions for BenMAP users. Clarification is needed in the manual for how BenMAP internally selects or adjusts CR functions to account for seasonal data. We note the absence of file headers for metadata in the air quality surface files and suggest such functionality be added. The 12km grid definition should be provided in the manual, the tool, and external references.

Panelist Ling Jin verified that the current version of BenMAP can incorporate externally created air quality surfaces on sub-domains of the 12 km grid to which the tool is currently limited (Figure A). We

express interest in standardized functionally or external code for re-gridding and processing air quality surfaces like the AQS data formatting script available on BenMap CE website.[12]

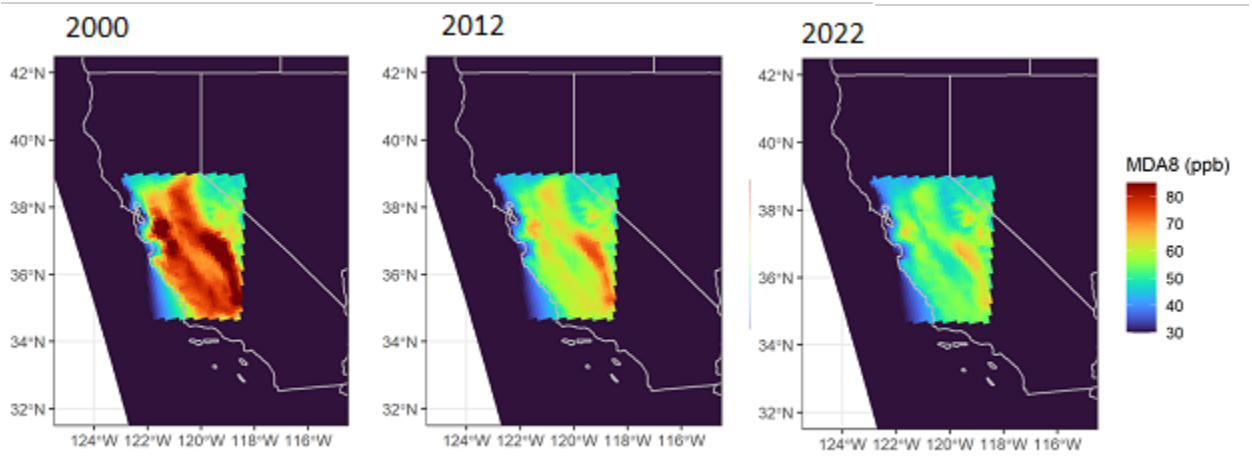



*Figure 1: Externally generated MDA8 Ozone surface uploaded on the 12 km grid for a central California sub-domain for three emissions years (source: Ling Jin)*

The air quality surfaces currently populated in the online version of BenMAP were not described or cited in the *User's Manual*, and no metadata about their policy cases or provenance were provided in the online tool, including in the available (currently empty) text boxes below them. **We recommend "info buttons" for metadata on air quality surfaces, links to describe their creation (emissions, meteorology, chemical transport model, etc.) and guidance on how to use each case** (Tier 2). The ability to share externally created air quality surfaces and analysis templates for public re-use in the data center is a common request. Performance benchmarks for modeled air quality surfaces are important guidance and metadata to prioritize in development.

Tools for using air quality monitoring data beyond monitor rollback are suggested, including monitor grid creation, quality assurance, the ability to fuse modeled and observed surfaces to reduce bias modeled inputs, and the spatial processing components included in BenMAP-CE and its offline scripts.

One major limitation that the *User's Manual* mentions is that BenMAP only resolves PM2.5 and ozone independently and does not necessarily resolve the full health and economic burdens of changes in air pollution considering the combined effects of both compounds or effects from ozone precursors, aerosol speciation, and other gases. **We recommend that EPA provide context or references on the importance of these limitations and for the Agency to consider functionality to estimate health impacts of changes to other pollutants and interaction terms for PM2.5 and ozone (Tier 2).** The latter reflect important ongoing scientific and model developments for regulatory and research applications.

[12] https://www.epa.gov/benmap

## 2.7. Charge Question 7: Technical Implementation and Transparency

*Please comment on the technical implementation and transparency of the tool's calculations and presentation of output results. Please include discussion on the extent to which:*

**2.7.a.** *results are correctly calculated, given the technical descriptions;*

The TSD and *User's Manual* generally did a good job documenting a very complex tool, however this question seems to ask whether the different versions of the BenMAP tool (desktop, web) implement the scientific procedures described in the TSD and *User's Manual* exactly as described. In a tool this complex, it is difficult or impossible to guarantee a "bug free" implementation and difficult to even evaluate by the review process undertaken by the SAB.

Instead, as discussed above in 2.6.c, BenMAP's development could be enhanced with a set of best practices to promote transparency and aid in identifying and correcting any errors that might exist. The BenMAP source code should be available for inspection by all users, ideally via a widely used repository such as GitHub. Additionally, a BenMAP Users Group, similar to the CMAQ users group would help in this regard. **As a Tier 1 recommendation, when EPA creates a new version of BenMAP (e.g., the web version) or makes updates to BenMAP, the EPA should run standard benchmarks to ensure that things are working as intended**. In particular, the EPA could reproduce with the web version of BenMAP some reference results from the desktop version. Benchmark results should be made available to users as output files that can be downloaded with summary tables and figures in appendices to the *User's Manual*. More documentation embedded in the user interface of the online tool would also help in this regard.

**More specifically, outputs should carry both text identifiers (e.g., county names) as well as Federal Information Processing System (FIPS) codes to make it easier for users to navigate (Tier 3). Model outputs should be stratified by demography, e.g., by race and ethnicity (Tier 3).**

**2.7.b***. documentation clearly describes how results are calculated;*

The documentation in the *Users' Manual* and TSD adequately describes how the results are calculated in general. Some of the ozone related health effects are not well documented. For example, the online version of BenMap computes ozone acute effects for D1hourmax (highest hourly value from 12:00 A.M. through 11:59 P.M.) even when the metric provided is D8hourMax (highest eight-hour average calculated between 12:00 A.M. and 11:59 P.M.). This could be an error or incorrect. It is not clear how the incidences are computed absent the corresponding AQ surface. About health endpoints for ozone, it is unclear how to adjust the CR function for annual versus warm season metrics. For mortality, the studies cited are all conducted for warm seasons indicating that the CR functions are only applicable to warm season metrics. However, online BenMAP returns outputs for ozone acute effects regardless of annual or warm season metrics provided in the air quality surface.

Additionally, as noted previously, while there is extensive documentation within the TSD and the *User's Manual*, the information provided during the user's experience within the model interface itself is very limited (see also response to Charge Question 6b).

**2.7.c.** *results are presented clearly and with adequate explanation;*

While the *Users' Manual* does a good job explaining how to run BenMAP, it lacks in its description of how to use the results and how to present results. This starts with the figures of the BenMAP results being un-readable in that the text is so fine that you cannot readily read the words. Zooming in on the text, at least with the pdf available, led to blurry letters as well. **We recommend that EPA make the text legible and then explain each field using the example of the text shown (Tier 2).** There is virtually no discussion of the results from the economic valuation and what each of the fields means. It would be useful to have the latitude and longitude of the grid results in the .csv file (or at least the option to provide them). Units should be provided in the output .csv file.

As noted elsewhere in this review, the explanation of the uncertainties needs to be more precise as it is not clear exactly how they are propagated from the grid-by-grid estimates to the aggregated results, and then to the estimated benefits.

**2.7.d.** *results are useful and appropriate for the intended scientific or policy purposes as documented; and*

We found the results useful and appropriate for the intended scientific and policy purposes as documented. Analyses at finer spatial resolutions (most commonly 1km) are supported by contemporary population data and air quality surfaces and necessary for many contemporary scientific applications and for environmental justice analyses. We would like to see guidance and references on the limitations of 12km and county scales for current policy purposes. We would also like to see guidance on how EPA applies results in cases of tradeoffs among different pollutants and tradeoffs and overlaps among different health endpoints. See also our response to 2.7.e below regarding presentation of uncertainties. Again the SAB encourages cooperation with the EPA's NEXUS team who have implemented environmental justice indictors into their web tools, including source proximity and demographic indicators, as well as additional exposures to climate impacts and air toxics that could enable a more holistic evaluation of multiple environmental stressors.

**2.7.e***. potential biases or uncertainties have been appropriately indicated and appropriate user notices are made available for practitioners applying the tool to their scientific or policy analyses. For reference, the "User Manual for BenMAP" describes the presentation of results [Chapter 10].*

**We recommend the model and its supporting documentation should more thoroughly address and explain the multiple sources of uncertainty in the analysis and its results (Tier 2).** There is clearly uncertainty at every step of the BenMAP modeling process: air quality data, population and demographic estimates, baseline health incidence, data, epidemiological CR relationships, and valuation functions. BenMAP does not provide means to include, estimate, or propagate uncertainties in the population, incidence, and air quality surface data, only for CR functions. Adding this functionality would be very useful. The Agency should ideally provide guidance and performance benchmarking for biases and uncertainties in each user-provided dataset and in net uncertainties in BenMAP health impacts assessments and valuation, and tools to calculate them. Net uncertainties are also important for both scientific and policy purposes to provide a signal to noise ratio for scientific phenomena and policy cases and to contextualize the magnitude of the response calculated. The

Monte Carlo approach discussed in our response to Charge Question 2.5.b would be a good way to develop a net uncertainty estimate. We specifically identified quantified margins of error in Census and American Community Survey population data and estimated uncertainties for baseline incidence that should be added to BenMAP as well as, highlighted a need for documentation on addressing uncertainties in population projections using Woods & Poole data and in projecting inflation and income growth.

**We recommend that BenMAP include descriptions and references for individual components of the analysis (Tier 2).** BenMAP quantifies the probability distribution for point estimates, reporting point estimate, mean, standard deviation, 2.5% and 97.5% (i.e., 95% confidence bounds). The documentation states that "BenMAP uses an N-point Latin Hypercube to represent the underlying distribution of β and to create a corresponding distribution of incidence changes in each population grid cell, where N is specified by you." However, in the tool and documentation, it is not clear where users can specify "N" for uncertainty analysis. The *User's Manual* discussed potential biases regarding the usage of VSL and COI for monetized health benefits, an area of ongoing debate. Users would benefit from additional guidance and flexibility in defining their own valuation functions and parameters. We also suggest that BenMAP include a description of potential biases and uncertainties related to modeling health and economic impacts at 12 km and county resolutions, and the impacts of extreme events.

## 2.8. Charge Question 8: User Interface

*Charge Question 8: In answering questions 8(a)-d(d), please consider the user interface, software platform and implementation, data formats, documentation and the user experience of the BenMAP software. Do these elements support the user's ability to:*

**2.8.a.** *make appropriate analytic choices;*

The offline documentation for the desktop version of BenMAP used for the TSD is comprehensive and impressive, but the online documentation links are very sparse. **We recommend that EPA provide a description of all model inputs used to create the model outputs through the online interface (Tier 1).** As an example, the air quality datasets provided in the online/cloud version of BenMAP need more description to guide the user. Abbreviations for model inputs and outputs also need to be clearly defined, and units for each variable need to be provided. All this information should be available by hovering a mouse over variable names in the online version. Further links to a summary discussion of model performance metrics would also be useful so that BenMAP users can judge the quality of the pre-defined exposure fields. More direct links from the BenMAP user interface to the supporting background information will enable the broader scientific community to use the BenMAP model, not just internal EPA experts.

**2.8.b.** *conduct a novel analysis;*

As discussed above, the current state of the cloud version of BenMAP is highly restrictive to the analyst's choices, significantly hampering their ability to conduct a novel analysis. As of this review, these restrictions apply to virtually each step of the online BenMAP's analytic process. For air quality inputs, the user seems to be restricted to a variety of pre-policy and post-policy air quality surfaces,

and the process of uploading user-specific air quality data is not clear, and the functionality to use air monitoring data is not yet incorporated into the model. Moreover, it does not seem possible to upload a geographic grid shape file that is not the preloaded counties or the CMAQ 12km grid. For both population and baseline health incidence inputs, the user cannot enter their own data and is restricted to pre-loaded U.S. population and incidence data. The analyst faces even more significant restrictions on their choice of health effects to evaluate and the health effects studies to estimate PM2.5- and ozone-attributable impacts. The user cannot select individual health effects, only subsets of all mortality or morbidity effects. Moreover, the user cannot input their own health impact functions and can only use those preloaded into the model. Finally, the analyst cannot pool together multiple health effect studies in their analysis. Generally, these restrictions on model input choices result in the online version of BenMAP user locked into an analysis path with few opportunities to customize and conduct a novel analysis.

In addition to restrictions in model inputs, there are potential novel analyses that either are not currently possible with BenMAP or not well documented within the software or its supporting documents. For example, the *User's Manual* states that in addition to modelling a small change in air quality resulting from a specific policy or action, another application of BenMAP is to assess the total health burden of current levels of air pollution. In the past, studies using BenMAP have done this exactly (e.g., Fann et al, 2012). However, there is no guidance in the *User's Manual* as to how to estimate the total health burden of air pollution.

Other limitations to the ability to conduct analysis are beyond the current functionality of BenMAP. BenMAP can only evaluate impacts of PM2.5 and ozone. Impacts of other pollutants, or sub-species of PM2.5, are outside the current scope of BenMAP, but expanding this functionality would further equip the analyst to assess attributable health impacts of air pollution. Other expansions of BenMAP functionality that could expand the ability to conduct novel analyses include incorporation of interactions between climate change and the health impacts of atmospheric pollutants and distributional analysis that incorporates consideration of equity and environmental justice across demographic subgroups, including indigenous tribes and other disproportionately burdened communities, in the estimation of health and economic impacts of air pollution.

**2.8.c.** *revisit a prior analysis to either replicate it or edit and rerun the prior analysis; and*

One somewhat-related current drawback of online BenMAP that hampers revisiting of a prior analysis is the fact that the user gets timed out of the program quickly, and in so doing seems to lose work in progress. Thus, after re-logging into the program, the user cannot return to the analysis that was in process. **We recommend that future versions of BenMAP fix the timing out problem (Tier 2)**. Other than this, we did not identify limitations with the BenMAP user's ability to revisit a prior analysis.

**2.8.d***. appropriately interpret and present results?*

For the most part, the current implementation does what it says it does, though it is very limited. The interface is easy to use though, **as a Tier 3 recommendation, we recommend links in the online BenMAP to the part of the *User's Manual* that covers that part of the program, i.e., a simple click to an explanation is enabled for each step where input or execution is optional or required from the**

**user, which includes information on how to utilize and interpret the results**. The data format is reasonably commonly used, though a bit painful to use. Indeed, a main limitation of the current BenMAP implementation is the rigidity of the input form (12 km grid following a specific CMAQ domain). While we did not pursue this further, it seems that it would be possible to input a data file that does not strictly align with the domain currently enabled, and the user would be unaware of such, including potentially using a different "origin corner." As noted elsewhere, more flexibility should be built into BenMAP, but also the way data are provided to BenMAP should be more flexible, particularly if it is going to be tied to CMAQ grid structures. It should be adapted to use the network Common Data Format (netCDF) files, given that is how CMAQ fields tend to be stored and used. One could readily develop a part to the user interface to use a CMAQ file.The approach should also be ready to utilize future CMAQ (and other model) formulations (e.g., when CMAQ starts using the Model for Prediction Across Scales (MPAS) system). As noted above, the documentation does lay out the inputs and how to run the model as implemented, though it is lacking when it describes the outputs, particularly for the valuations.

### 2.9. Charge Question 9: Future Peer Review

*Charge Question 9: Please comment on what types of changes the SAB believes warrant subsequent peer review of the BenMAP software.*

Many changes and updates to BenMAP's analytic choices follow changes to the EPA's PM and ozone Integrated Science Assessments (ISAs). Because the ISAs are peer reviewed, additional peer review for updates to BenMAP that reflect the latest science from the ISAs would be redundant and unnecessary. For example, changing the health effects resulting from PM2.5 and ozone exposure that meet the causality threshold required for inclusion in EPA's core suite of PM2.5- and ozone-attributable health effects, or the preferred health effects studies to quantify these attributable impacts in BenMAP, would not need peer review.

Other changes to the software, however, that expand or change the capabilities or functionality of BenMAP should undergo peer review. Below we list several such changes that would qualify as functional changes that would warrant peer review, including:

- addition of new pollutants, including speciation of PM2.5, that BenMAP could evaluate;
- significant changes to spatial or temporal resolution for BenMAP inputs and/or outputs;
- significant changes in valuation approaches, such as shifting from the current cost-of-illness based-studies for valuing morbidity impacts to more comprehensive willingness-to-pay approaches;
- changes to the model such that interactions between climate change and health impacts of atmospheric pollutants are taken into account;
- incorporation of distributional analysis that would present opportunities for new applications of BenMAP; and
- substantive changes to how BenMAP treats uncertainty.

As the BenMAP model evolves in the above ways, or in ways not anticipated here, peer review would not necessarily require review of the entire model. Instead, peer review could be more focused on the

specific change(s) and done by a smaller group, potentially with the aid of a contractor that could conduct specific calculations determined by the peer-review group.

Finally, in addition to peer review for major changes to the model, we believe that periodic review, perhaps every five years, would be warranted to ensure that the model is up to date and reflects the latest science and economics. This is especially important if BenMAP is to continue to be a central model used by the EPA in benefit-cost analyses for air regulations. Such periodic reviews could include review of changes to future health incidence projections, population projections, CR functions, valuation functions, and other features of the BenMAP model.

### 2.10. Charge Question 10: Emerging Trends

*Charge Question 10: Please discuss any emerging trends in software development, computer science, or data management technologies that should be considered or prioritized, for future upgrades to the BenMAP software or similar tools.*

BenMAP is useful both for its algorithmic pipeline and its current parameterizations and input data (population, incidence rate, CR function and valuation functions, etc.). The development already considers architecture that separates graphical user interfaces (GUI) and scientific computation, client and server, and can support more flexibility for the research communities to apply BenMAP's algorithmic pipeline to user-defined input data and scientific, computational, analytical, data management, and visualization extensions. Additional functionalities to handle additional and emerging datasets such as satellite data, high resolution spatial and temporal data from low-cost sensor networks, mobile sensors, and traffic data, could enhance data quality and greatly expand the user community. The near-term availability of geostationary satellite remote sensing data for ozone and its precursors from other federal agencies at high spatial and temporal resolution was highlighted as an important trend to incorporate in future releases.

We note that the user community may likely prefer a non-cloud desktop or software library version of BenMAP available that replicates and extends functionality of the official cloud version. **To expand on our recommendation in 2.6.c, we recommend that EPA package BenMAP source code and components such as preprocessing modules, core computation modules, and post processing modules, so that the research community can grow ecosystems around it (Tier 2**). BenMAP would benefit from a centralized data sharing and code platform for cross validation and innovation to support research applications and community health impacts assessment tool research and development following the CMAS center community modelling approach while agency-led development is constrained**. More specifically, we recommend an R or python implementation be available and able to be downloaded from community software repositories such as anaconda[13] and r-cran[14] before availability and support for the desktop version of BenMAP is discontinued (Tier 2).**

**We also recommend unit tests for all algorithms (Tier 2).** We encourage clear standards for unit tests, code organization, code comments, and systematic dependencies with clarity and consistency across the user manual, code documentation and markup, and GitHub repositories.

[13] https://anaconda.org/anaconda/repo

[14] https://cran.r-project.org/web/packages/available_packages_by_name.html

Coordination with development in the Models-3 ecosystem and upstream models would support BenMAP or similar tools in incorporating emerging trends in software development, computer science, and data management. For example, development of variable-scale, nesting, and flexible grids is ongoing in the chemical transport models used to create air quality surfaces for BenMAP (CMAQ, WRF-Chem, Community Earth System Model (CESM)). When CMAQ moves to a new grid structure, the cloud based BenMAP should be ready to use the fields developed using the new grid system. The adjoint and linear tangent of the CMAQ model can support more sophisticated sensitivity analyses and we expect this to be a growing trend in scientific and policy applications of BenMAP and similar tools.

Ensuring data meet principles of findability, accessibility, interoperability, and reusability (FAIR; Wilkinson et al., 2016) is an emerging data management and sharing priority among the scientific communities that apply BenMAP and contribute to its many inputs, and a contemporary priority across several federal agencies, notably NIH and NIST. **We recommend all future development of BenMAP pursue FAIR standards in all datasets and choices in data, metadata, log files, and code (Tier 1).** Two examples of changes that would improve alignment with FAIR standards are standardizing the log file format to include all input data sources and parameter choices and conform to a machine-readable standard and providing documentation for the BenMAP database schema and code examples in Python and R detailing how to connect to the database and extract data from it. Replicability in the current version of the tool can be greatly enhanced with minimal development by supporting public sharing of user-added data and analysis templates. Models-3 has long conformed to FAIR principles in standardized formats, and this presents an effective initial area for coordination. Machine-readable data and metadata will support incorporation of BenMAP results into sophisticated visualization suites (e.g., Tableau), automated evidence maps, and re-use of scientific and policy applications of BenMAP.

We highlight machine learning tools and large language models as important emerging trends to consider across code development, testing, documentation, analysis, sensitivity evaluation, optimization, replication. Making BenMAP data FAIR will ensure compatibility with these trends. Among currently available tools, we identified potential value in applying GitHub Copilot for coding suggestions, Mintlify to generate documentation from code, Deepcode to identify bugs and security vulnerabilities, and large language models to convert code between programming languages and standards and convert documentation to other written languages.

**REFERENCES**


Agency for Healthcare Research and Quality. (n.d.). Medical Expenditure Panel Survey. Retrieved from https://meps.ahrq.gov/mepsweb/index.jsp

Aguilar-Gomez S, Dwyer H, Graff-Zivin J and Neidell M. (2022). This is Air: The "Non-Health" Effects of Air Pollution. *Annual Review of Resource Economics*, 403-425.

Ahn S and Becker BJ. (2011). Incorporating quality scores in meta-analysis. *Journal of Educational and Behavioral Statistics, 36*(5), 555-585.

Akoumianakis I, Zagaliotis A, Konstantaraki M and Filippatos TD. (2023, May 16). GLP-1 analogs and regional adiposity: A systematic review and meta-analysis. *Obesity Review*, e13574. doi:10.1111/obr.13574

Alberini A, Loomes G, Scasny M and Bateman I. (2010). *Valuation of Environment-Related Health Risks for Children.* Organization for Economic Cooperation and Development (OECD). OECD Publishing. Retrieved from https://www.oecd-ilibrary.org/environment/valuation-of-environment-related-health-risks-for-children_9789264038042-en

Alzheimer's Association. (2023). *Alzheimer's Disease Facts and Figures*. Retrieved from Alzheimer's Disease Facts and Figures: https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf

Aucejo EB and Romano TF. (2016). Assessing the Effect of School Days and Absences on Test Score Performance. *Economics of Education Review, 55*, 70-87.

Avanth CV, Rutherford C, Rosenfeld EB, Brandt JS, Graham H, Kostis WJ and Keyes KM. (2023, May 22). Epidemiologic trends and risk factors associated with the decline in mortality from coronary heart disease in the United States, 1990 - 2019. *American Heart Journal*, 46-55. doi:10.1016/j.ahj.2023.05.006

Benjamin DJ. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics vol 2*, DOI: 10.1016/bs.hesbe.2018.11.002.

Bollinger CR and Hirsch BT. (2006). Match bias from earnings imputation in the Current Population Survey: The case of imperfect matching. *Journal of Labor Economics, 24*(3), 483-519.

Borenstein M, Hedges LV, Higgins JPT, Rothstein H. (2010). A Basic Introduction to Fixed-effect and Random-effects Models for Meta Analysis. *Research Synthesis Methods*, 97-111.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. (2021). *Introduction to Meta-Analysis, 2nd ed.* Hoboken, NJ: John Wiley & Sons.

Bowlands BJ and Beghin JC. (2001). Robust estimates of the value of statistical life for developing economies. *Journal of Policy Modeling, 23*(4), 385-396.

Brook JR, Dann TF and Burnett RT. (2011). The Relationship Among TSP, PM10, PM2.5, and Inorganic Constituents of Atmospheric Participate Matter at Multiple Canadian Locations. *Journal of Air and Waste Management Association, 47*(12-19). Retrieved from https://www.tandfonline.com/doi/abs/10.1080/10473289.1997.10464407

Burnett R, Chen H, Szyszkowicz M +50. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences, 115*(38), 9592-9597. doi:https://doi.org/10.1073/pnas.1803222115

Calderon MA, Alves B, Jacobson B, Hurwitz A, Sheikh A and Durham S. (2007, January 24). Allergen Injection Immunotherapy for Seasonal Allergic Rhinitis. *Cochrane Database of Systematic Reviews*. doi:10.1002/14651858.CD001936.pub2

Cameron TA. (2001). *Review of the Draft Analytical Plan for EPA's Second Prospetive Analysis--Benefits and Costs of the Clean Air Act 1990 - 2020.* Advisory Council on Clean Air Compliance Analysis. EPA Science Advisory Board. Retrieved from https://council.epa.gov/ords/sab/f?p=104:12:12984227144898

Cameron TA and DeShazo JR. (2013). Demand for Health Risk Reductions. *Journal of Environmental Economics and Management, 65*, 87-109.

Cameron TA, DeShazo JR, Johnson E. (2009). Willingness to Pay for Health Risk Reductions: Differences by Type of Illness. *unpublished manuscript*.

Cameron TA, DeShazo JR, Johnson EH. (2011). Scenario adjustment in stated preference research. *Journal of Choice Modelling, 4*(1), 9-43. Retrieved from https://www.sciencedirect.com/science/article/pii/S1755534513700174

Center for Disease Control. (2018). *CDC Social Vulnerability Index 2018 Documentation.* Retrieved from https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/pdf/SVI2018Documentation-H.pdf

Center for Disease Control. (n.d.). *Chronic Disease and Health Promotion Data & Indicators*. Retrieved from https://chronicdata.cdc.gov/browse?category=500_Cities=%26+Places

Center for Medicare & Medicaid Services. (n.d.). *Chronic Conditions Data Warehouse*. Retrieved from Chronic Conditions Data Warehouse: https://www2.ccwdata.org/web/guest/home/

Chakraborty S, Dey T, et al. (2022). A Spatiotemporal Analytical Outlook of the Exposure to Air Pollution and COVID-19 Mortality in the USA. *Journal of Agricultural, Biological and Environmental Statistics*, 419-439. Retrieved from https://pubmed.ncbi.nlm.nih.gov/35106052/

Chandra A, Coile C and Mommaerts C. (2023 forthcoming). What Can Economics Say About Alzheimer's Disease? *Journal of Economic Literature*. doi:10.3386/w27760

Chen Z, Sidell MA, et al. (2022). Ambient Air Pollutant Exposures and COVID-19 Severity and Mortality in a Cohort of Patients with COVID-19 in Southern California. *American Journal of Respiratory and Critical Care Medicine*, 440-448. Retrieved from https://pubmed.ncbi.nlm.nih.gov/35537137/

Chetty R, Friedman JN and Rockoff JE. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 2633-2679. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.104.9.2633

Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: extending omitted variable bias. *Journal of the royal Statistical Society, 82*(1), 39-67.

Colmer J and Voorheis J. (2020). *The Grandkids Aren't Alright: the Intergenerational Effects of Prenatal Pollution Exposure.* Census Working Papers, U.S. Census Bureau. Working paper Number CES-20-36. Retrieved from https://www.census.gov/library/working-papers/2020/adrm/CES-WP-20-36.html

Cooper H, Hedges LV, Valentine JC. (2019). *The Handbook of Research Syntehsis and Meta-Analysis, 3rd ed.* New York, NY: Russel Sage Foundation.

Cordray DS, Morphy P. (2009). The Handbook of Research Synthesis and Meta-Analyses. *Research Synthesis and Public Policy*, 473-494.

Costa DL and Kahn ME. (2004). Changes in the Value of Life, 1940-1980. *Journal of Risk and Uncertainty, 29*, 159-180.

Council of Economic Advisors. (2022). *The Annual Report of the Council of Economic Advisors.* Retrieved from https://www.govinfo.gov/content/pkg/ERP-2022/pdf/ERP-2022.pdf

Cropper ML and Krupnick AJ. (1999). The social costs of chronic heart and lung disease. In *Valuing Environmental Benefits: Selected Essays of Maureen Cropper.* Edward Elgar. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4242a9cc58bbb47e88e31e6548e4c106319e16a3

Crouse DL, Erickson AC, Christidis T, Pinault L + 10. (2020). Evaluating the Sensitivity of PM2.5-Mortality Associations to the Spatial and Temporal Scale of Exosure Assessment. *Epidemiology, 31*(2), 168-176. Retrieved from https://journals.lww.com/epidem/Abstract/2020/03000/Evaluating_the_Sensitivity_of_PM2_5_Mortality.3.aspx

Currie J, Neidell M and Schmeider JF. (2009). Air Pollution and Infant Health: Lessons from New Jersey. *Journal of Health Economics, 28*(3), 688-703.

Dellink R, Chateau J, Lanzi E and Magne B. (2017, January). Long-term economic growth projections in the Shared Socioeconomic Pathways. *Global Environmental Change, 42*, 200-214. Retrieved from https://www.sciencedirect.com/science/article/pii/S0959378015000837?via%3Dihub

Depsky N, Cushing L and Morello-Frosch T. (2022). High-resolution gridded estimates of population sociodemographics from the 2020 census in California. *PLoS ONE, 17*(7). Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270746

Deryugina T, Heutel G, Miller NH, Molitor D and Reif J. (2019). The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. *American Economic Review, 109*, 4178-4219. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.20180279

Deschenes O, Greenstone M and Shapiro JS. (2017). Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program. *American Economic Review, 107*, 2958-89.

Di Q, Amini H, Shi L, Kloog I +9. (2019). An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. *Environment International, 130*, 104909.

Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, Dominici F. (2017a). Association of Short-term Exposure to Air Pollution With Mortality in Older Adults. *Journal of the American Medical Association, 318*(24), 2446. Retrieved from https://pubmed.ncbi.nlm.nih.gov/29279932/

Di Q, Wang Y, Zanobetti A, Koutrakis P, Choirat C, Dominici F and Schwartz JD. (2017b). Air Pollution and Mortality in the Medicare Population. *New England Journal of Medicine, 376*(26), 2513-22. doi:10.1056/NEJMoa1702747

Dickie M and Messman VL. (2004). Parental altruism and the value of avoiding acute illness: are kids worth more than parents? *Journal of Environmental Economics and Management, 48*(3), 1146-1174.

Doucouliagos H, Stanley TD and Viscusi WK. (2014). Publication selection and the income elasticity of the value of a statistica life. *Journal of Health Economics, 33*, 67-75.

Drummond M, Helfand M, and Mullins CD, Editors. (2012, September-October). Medical Decision Making Special Issue: Recommendations of the ISPOR-SMDM Joint Modeling Good Research Practices Task Force. In H. M. Drummond M (Ed.)., (pp. 653-743). doi:https://doi.org/10.1177/0272989X12458665

Duinker PN and Greig LA. (2007). Scenario analysis in environmental impact assessment: Improving explorations of the future. *Environmental Impact Assessment Review, 27*(3), 206-219.

Ebenstein A, Lavy B and Roth S. (2016). The Long-Run Economic Consequences of High Stakes Examinations: Evidence from Transitory Variation in Pollution. *American Economic Journal: Applied Economics, 8*(4), 36-65.

Eum KD, Kazemiparkouhi F, Wang B, Manjourides J, Pun V, Pavlu V and Suh H. (2019, Mar). Long-term NO2 exposures and cause-specific mortality in American older Adults. *Environment International*, 10-15. doi:10.1016/j.envint.2018.12.060

European Chemicals Agency. (2016). *Valuing Selected Health Impacts of Chemicals.* Retrieved from https://echa.europa.eu/documents/10162/13630/echa_review_wtp_en.pdf/dfc3f035-7aa8-4c7b-90ad-4f7d01b6e0bc

Fabian MP, Adamkiewicz G, Stout NK, Sandel M and Levy JI. (2014, January). A simulation model of building intervention impacts on indoor environmental quality, pediatric asthma, and costs. *Journal of Allergy and Clinical Immunology, 133*(1), 77-84. doi:10.1016/j.jaci.2013.06.003

Fann N, Lamson AD, Anenberg S, Wesson K, Risley D and Hubbell BJ. (2012). Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone. *Risk Analysis, 32*(1), 81-95. doi:10.1111/j.1539-6924.2011.01630.x

Freeman AM. (2006). Valuing Environmental Health Effects - An Economic Perspective. *Environmental and Resource Economics, 34*, 347-363.

Garcia E, Johnston J, McConnell R, Palinkas L and Eckel SP. (2023). California's early transition to electric vehicles: Observed health and air quality co-benefits. *The Science of the Total Environment, 867*, 161761. Retrieved from https://pubmed.ncbi.nlm.nih.gov/36739036/

Garland A, Jong SH, Stepner M +6. (2019). Effects of Cardiovascular and Cerebrovascular Health Events on Work and Earnings: A Population-Based Retrospective Cohort Study. *Canadian Medical Association Journal, 191*, E3-E10. Retrieved from https://pubmed.ncbi.nlm.nih.gov/30617227/

Gayer T, Hamilton JT and Viscusi WK. (2000). Private Values of Risk Tradeoffs at Superfund Sites: Housing Market Evidence on Learning about Risk. *Review of Economics and Statistics, 82*(3), 439-51.

Gerking S and Dickie M. (2013). Valuing Reductions in Environmental Risks to Children's Health. *Annual Review of Resource Economics, 5*, 245-260. doi:https://doi.org/10.1146/annurev-resource-091912-151940

Gharibvand L, ShavlikD, Ghamsary M, Beeson WL, Soret S, Knutsen R and Knutsen SF. (2017, March). The Association between Ambient Fine Particulate Air Pollution and Lung Cancer Incidence: Results from the AHSMOG-2 Study. *Environmental Health Perspectives, 125*(3), 378-385. doi:10.1289/EHP124

Gilliland FD, Berhane K, Rappoport EB +8. (2001). The effects of ambient air pollution on school absenteeism due to respiratory illnesses. *Epidemiology*, 43-54.

Goldsmith L, Bell ML. (2022). Queering Environmental Justice: Unequal Environmental Health Burden on the LGBTQ+ Community. *Am J Public Health*, 79-87.

Graff Zivin J, Neidell M. (2013). Environment, Health, and Human Capital. *Journal of Economic Literature, 51*(3), 689-730.

Graff-Zivin J and Neidell M. (2012). The Impact of Pollution on Worker Productivity. *American Economic Review, 102*(7), 3652-73.

Hamer D and Howson D. (1978). Bibliography on electroanalgesia. *Physical Therapy*, 1485-92.

Hammitt JK. (2008). *Characterizing Uncertainty in Particulate Matter Benefits Using Expert Elicitation.* Advisory Council on Clean Air Compliance Analysis. EPA Science Advisory Board Staff Office. Retrieved from https://council.epa.gov/ords/sab/f?p=104:12:12984227144898

Hammitt JK. (2017a). Valuing Non-Fatal Health Risks. *Revue économique, 68*(3), 335-356. Retrieved from https://www.jstor.org/stable/90013457

Hammitt JK. (2017b). Extrapolating the Value Per Statistical Life Between Populations: Theoretical Implications. *Journal of Benefit-Cost Analysis, 8*(2), 215-225. Retrieved from https://www.cambridge.org/core/journals/journal-of-benefit-cost-analysis/article/extrapolating-the-value-per-statistical-life-between-populations-theoretical-implications/706C4D57A37B841D5DF365438D10C44B

Hammitt JK and Haninger K. (2017). Valuing nonfatal health risk as a function of illness severity and duration: benefit transfer using QALYs. *Journal of Environmental Economics and Management, 8*, 17-38.

Hammitt JK and Robinson LA. (2011). The Income Elasticity of the Value per Statistical Life: Transferring Estimates between High and Low Income Populations. *Journal of Benefit-Cost Analysis, 2*(1), 1-29. doi:10.2202/2152-2812.1009

Hanmer J, Lawrence WF, Anderson JP, Kaplan RB and Fryback DG. (2006, July-August). Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. *Medical Decision Making, 26*(4), 391-400. doi:10.1177/0272989X06290497

Hardin FM, Eskander PN and Franzese C. (2021). Cost-effective Analysis of Subcutaneous vs Sublingual Immunotherapy from the Payor's Perspective. *OTO Open, 5*(4), 1-6. Retrieved from https://pubmed.ncbi.nlm.nih.gov/34723051/

Hauer M. (2018). *Population projections for all U.S. counties by age, sex, and race controlled to the Shared Socioeconomic Pathways*. doi:10.17605/OSF.IO/9YNFC

Hauer M. (2019). Population projections for US counties by age, sex and race controlled to shared socioeconomic pathway. *Scientific data, 6*(1), 1-15.

Hill W, Lim EL, Weeden CE, +69. (2023). Lung adenocarcinoma promotion by air pollutants. *Nature, 616*, 159-167. Retrieved from https://www.nature.com/articles/s41586-023-05874-3

Hirsch BT and Schumacher EJ. (2004). Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics, 22*(3), 689-722.

Holub F, Hospido L, Wagner UJ. (2020). Urban Air Pollution and Sick Leaves: Evidence from Social Security Data. *Social Science Research Network*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3572565

Hunt A and Ferguson J. (2010). *A review of recent policy-relevant findings from the environmental health literature.* Organization for Economic Cooperation and Development. Retrieved from https://one.oecd.org/document/ENV/EPOC/WPNEP(2009)9/FINAL/en/pdf

Hunt A, Ferguson J, Hurley F & Searl A. (2016). *Social Costs of Morbidity Impacts of Air Pollution.* Organization for Economic Cooperation and Development. OECD Publishing. Retrieved from https://ideas.repec.org/p/oec/envaaa/99-en.html

Institute of Medicine. (2006). *Valuing Health for Regulatory Cost-Effectiveness Analysis.* Washington, DC: The National Academies Press. doi:https://doi.org/10.17226/11534

Interagency Working Group Paper. (2021). *Technical Support Document: Social Cost of Carbon, Methane and Nitrous Oxide.* White House. Retrieved from https://www.whitehouse.gov/wp-content/uploads/2021/02/TechnicalSupportDocument_SocialCostofCarbonMethaneNitrousOxide.pdf

Isen A, Rossein-Slater M, Walker WR. (2017). Every Breath You Take, Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970. *Journal of Political Economy, 125*(3). Retrieved from https://www.journals.uchicago.edu/doi/10.1086/691465

Jak S. (2015). *Meta-Analytic Structural Equation Modeling.* Switzerland: Springer.

Jak S, Cheung MWL. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods, 25*(4), 430-455.

Jak S, Li H, Kolbe L, de Jonge H, Cheung MWL. (2021). Meta-analytic structural equation modeling made easy: a tutorial and web application for one-stage MASEM. *Research Synthesis Methods, 12*, 590-606.

Johnson BH, Bonafedes MM and Watson C. (2016). Short and Longer-Term Health Care Resource Utilization and Costs Associated with Acute Ischemic Stroke. *ClinicoEconomics and Outcomes Research, 8*, 53-61.

Johnston RJ, Boyle KJ, Adamowicz W, Bennett J, Brouwer R, Cameron TA, Hanemann WM, Hanley N, Ryan M, Scarpa R, Tourangeau R and Vossler CA. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists, 4*(2), 319-405.

Kazemiparkouhi F, Eum KD, Wang B, Manjourides J, Suh HH. (2020, July). Long-term ozone exposures and cause-specific mortality in a US Medicare cohort. *J Expo Sci Environmental Epidemiology, 30*(4), 650-658.

Kenkel D. (2006). WTP and QALY-Based Approaches to Valuing Health for Policy: Common Ground and Disputed Territory. *Environmental and Resource Economics, 34*, 419-437.

Kennedy P. (2008). *A Guide to Econometrics 6th ed.* Malden, MA: Blackwell Publishing.

Kepes S, Wang W, Cortina JM. (2023). Heterogeneity in meta-analytic effect sizes: an assessment of the current state of the literature. *Organizational Research Methods*, DOI: 10.177/10944281231169942.

Kioumourtzoglou M, Schwartz JD, Weisskopf MG, Melly SJ, Wang Y, Dominici F and Zanobetti A. (2016, Jan). Long-term PM2.5 Exposure and Neurological Hospital Admissions in the Northeastern United States. *Environmental Health Perspectives*. Retrieved from https://ehp.niehs.nih.gov/doi/10.1289/ehp.1408973

Kleckner N and Neumann J. (1999). *Recommended Approach to Adjusting WTP.* Memorandum to Jim Democker, US EPA.

Klose T. (2003). A Utility-theoretic model for QALYs and Willingness to Pay. *Health Economics, 12*, 17-31.

Klose T. (2003). A Utility-theoretic model for QALYs and Willingnesss to Pay. *Health Economics, 12*, 17-31.

Kniesner TJ, Viscusi WK and Ziliak JP. (2010). Policy relevant heterogeneity in the value of statistical life: New evidence from panel data quantile regressions. *Journal of Risk and Uncertainty, 40*, 15-31.

Kones R and Rumana U. (2017, May). Cardiometabolic diseases of civilization: history and maturation of an evolving global threat. An update and call to action. *Annals of Medicine, 49*(3), 260-274.

Leggett CG. (2002). Environmental valuation with imperfect information: the case of the random utility model. *Environmental and Resource Economics, 23*, 343-355.

Lepeule J, Laden F, Dockery D and Schwartz J. (2012). Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974-2009. *Environmental Health Perspectives, 120*(7), 965-970.

Levy JI, Fabian MP, Peters JL. (2014, Jan 28). Community-Wide Health Risk Assessment Using Geographically Resolved Demographic Data: A Synthetic Population Approach. (T. YK, Ed.) *PLoS One*, e87144.

Lillard L, Smith JP and Welch F. (1986). What do we really know about wages? The importance of nonreporting and census imputation. *Journal of Political Economy, 95*(3, Part 1), 489-506.

Limaye VS, Vargo J, Harkey M, Holloway T and Patz JA. (2018). Climate Change and Heat-Related Excess Mortality in the Eastern USA. *Ecohealth, 15*(3), 485-496. doi:10.1007/s10393-018-1363-0

Little RJA and Rubin DB. (2014). *Incomplete Data.* John Wiley & Sons, Inc.

Liu J, Lee M and Gershenson S. (2021). The short and long-run impacts of secondary school absences. *Journal of Public Economics, 199*, 104441.

Liu JT, Hammitt JK and Liu JL. (1997). Estimated hedonic wage function and value of life in a developing country. *Economics Letters, 57*(3), 353-58.

Ludwig L and Neumann J. (2012, March 30). *Updating Income Elasticity Estimates in EPA's BenMAP Air Pollution Benefits Estimation System.* Retrieved from Memorandum to EPA Office of Air and Radiation: https://www3.epa.gov/ttnecas1/regdata/Benefits/Income%20Elasticity%20Update_Recommendations_noAppD.pdf

Manski CF. (2008). *Identification for Prediction and Decision.* Cambridge, MA: Harvard University Press.

Manski CF. (2010). Interpreting and combining heterogeneous survey forecasts. In M. Clements, & D. Henry, *Oxford Handbook on Economic Forecasting.*

Manski CF. (2020). Toward credible patient-centered meta-analysis. *Epidemiology, 31*(3), 345-352.

Marian B, Ying Y, Chen Z, Lurmann F, Li K, Gilliland F, Eckel SP and Garcia E. (2022). Independent associations of short- and long-term air pollution exposure with COVID-19 mortality among Californians. *Environmental Advances, 9*. doi:10.1016/j.envadv.2022.100280

Markowitz SB, Levin SM, Miller A, Morabia A. (2013, July 1). Asbestos, asbestosis, smoking, and lung cancer. New findings from the North American insulator cohort. *American Journal of Respiratory and Critical Care Medicine, 188*(1). doi:10.1164/rccm.201302-0257OC

Masterman C and Viscusi WK. (n.d.). The Income Elasticity of Global Values of a Statistical Life: Stated Preference Evidence. *Journal of Benefit-Cost Analysis, 9*(3), 407-434.

Menichini F and Mudu P. (2010). Drug Consumption and Air Pollution: An Overview. *Pharmacoepidemiology and Drug Safety*, 1300-15. Retrieved from https://pubmed.ncbi.nlm.nih.gov/20927798/

Miller TR. (2000). Variations between countries in values of statistical life. *Journal of Transport Economics and Policy*, 169-188.

Mrozek JR and Taylor LO. (2002). What determines the value of life? A meta-analysis. *Journal of Policy Analysis and Management, 21*(2).

Mu F, Hurley D, Betts K, Messali AJ, Paschoalin M, Kelley C and Wu EQ. (2017). Real-world costs of ischemic stroke by discharge status. *Current Medical Research and Opinion, 33*(2), 371-378.

Narain U and Sall C. (2016). *Methodology for valuing the health impacts of air pollution : discussion of challenges and proposed solutions.* Retrieved from World Bank Understanding Poverty web site: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/832141466999681767/methodology-for-valuing-the-health-impacts-of-air-pollution-discussion-of-challenges-and-proposed-solutions

Neumann PJ and others (eds). (2016). *Cost-Effectiveness in Health and Medicine, 2nd edn.* Oxford Academic Press. doi:https://doi.org/10.1093/acprof:oso/9780190492939.001.0001,

New York City Health Department. (n.d.). *Environment and Health Data Portal*. Retrieved from https://a816-dohbesp.nyc.gov/IndicatorPublic/beta/

Newbold SC. (2011). Valuing health risk changes using a life-cycle consumption framework. *NCEE Working Paper Series*, #11-03.

Nimdet K, Chalyakunapruk N, Vichansavakul K and Ngorsuraches S. (2015). A systematic review of studies eliciting willingness-to-pay per quality-adjusted life year: does it justify CE threshold? *PLOS One, 10*(4), e0122760. doi:10.1371/journal.pone.0122760

Nuclear Regulatory Commission. (2023, February 24). *Nuclear Regulatory Commission Policy Issue: Appendices K and L to NUREG/BR-0058, Revision 5.* Retrieved from https://www.nrc.gov/docs/ML2227/ML22273A170.pdf

Nurmatov U, Dhami S +31. (2017). Allergen immunotherapy for allergic rhinoconjunctivitis: a systematic overview of systematic reviews. *Clinical Translational Allergy, 7*(24), 1-16. Retrieved from https://pubmed.ncbi.nlm.nih.gov/28794855/

Office of Management and Budget. (2003). Circular A-4: Regulatory Analysis. Retrieved from https://obamawhitehouse.archives.gov/omb/circulars_a004_a-4/

O'Neill BC, Carter TR, Ebi K, Harrison PA +16. (2020, November 25). Achievements and needs for the climate change scenario framework. *Nature Climate Change*, 1074-1084. Retrieved from https://www.nature.com/articles/s41558-020-00952-0

Onukwugha E, McRae J, Kravetz A, Varga S, Khairnar R and Mullins CD. (2016). Cost-of-Illness Studies: An Updated Review of Current Methods. *Pharmacoeconomics, 34*(1), 43-58. doi:10.1007/s40273-015-0325-4

Organization for Economic Cooperation and Development. (2023). *Surveys on Willingness-to-Pay to Avoid Negative Chemicals-Related Health Impacts (SWACHE).* Retrieved from https://www.oecd.org/environment/tools-evaluation/costs-benefits-chemicals-regulation.htm

Ostro BC and Cameron TA. (2004). *Advisory on Plans for Health Effects Analysis in the Analytical Plan for EPA's Second Prospective Analysis - Benefits and Costs of the Clean Air Act, 1990 - 2020.* Advisory Council on Clean Air Compliance Analysis. EPA Science Advisory Board Staff Office. Retrieved from https://council.epa.gov/ords/sab/f?p=104:12:12984227144898

O'Sullivan AK, Rubin J, Nyambose J +3. (2011). Cost estimation of cardiovascular disease events in the U.S. *Pharmacoeconomics, 29*(8), 693-704. Retrieved from https://link.springer.com/article/10.2165/11584620-000000000-00000

Paolella DA, Tessum CW, Adams PJ, Apte JS, Chambliss S, Hill J, Muller NZ and Marshall JD. (2018). Effect of model spatial resolution on estimates of fine particulate matter exposure and exposure disparities in the U.S. *Environmental Science & Technology Letters*, 436-441. Retrieved from chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.cmu.edu/ceic/assets/docs/publications/published-papers/2017-and-2018/paolella-et-al-2018.pdf

Parry I, Black MS and Vernon N. (2021). *Still not getting energy prices right: a global and country update of fossil fuel subsidies.* International Monetary Fund. Retrieved from https://www.imf.org/en/Publications/WP/Issues/2021/09/23/Still-Not-Getting-Energy-Prices-Right-A-Global-and-Country-Update-of-Fossil-Fuel-Subsidies-466004

Patterson RF and Harley RA. (2019). Effects of freeway rerouting and boulevard replacement on air pollution exposure and neighborhood attributes. *International Journal of Environmental Research and Public Health*, 4072. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6862437/

Petkova EP, Horton RM, Bader DA and Kinney PL. (2013). Projected Heat-Related Mortality in the U.S. Urban Northeast. *International Journal on Environmental Research and Public Health, 10*(12), 6734-6747. doi:10.3390/ijerph10126734

Rahman MM, McConnell R +10. (2022). The Effects of Coexposure to Extremes of Heat and Particulate Air Pollution on Mortality in California: Implications for Climate Change. *American Journal of Respiratory and Critical Care Medicine*, 1117-1127. Retrieved from https://pubmed.ncbi.nlm.nih.gov/35727303/

Reid W. (1996). Willingness to pay and cost of illness for changes in health capital depreciation. *Econometrics and Health Economics, 5*, 447-468.

Reither E, Olshansky S, Yang Y. (2011). New forecasting methodology indicates more disease and earlier mortality ahead for today's younger Americans. *Health Affairs, 30*(8), 1562-1568.

Riahi K, Van Vuuren DP, Kriegler E +43. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change, 42*, 153-168. Retrieved from https://www.sciencedirect.com/science/article/pii/S0959378016300681

Robinson LA and Hammitt JK. (2013). Skills of the trade: valuing health risk reductions in benefit-cost anlaysis. *Journal of Benefit-Cost Analysis, 4*(1), 107-130.

Robinson LA and Hammitt JK. (2015). *The Effect of Income on the Value of Mortality and Morbidity Risk Reductions.* Retrieved from https://sab.epa.gov/ords/sab/f?p=114:0:14071292186454:APPLICATION_PROCESS=AGENCY_REVIEW:::AR_ID:2438).

Robinson LA, Hammitt JK, and O'Keefe L. (2019b). Valuing mortality risk reductions in global benefit-cost analysis. *Journal of Benefit Cost Analysis, 10*(S1), 15-50.

Robinson LA, Raich WJ, Hammitt JK and O'Keefe L. (2019a, Summer). Valuing Children's Fatality Risk Reductions. *Journal of Benefit-Cost Analysis, 10*(2). doi: https://doi.org/10.1017/bca.2019.10

Roth S. (2018). The Effect of Indoor Air Pollution on Cognitive Performance: Evidence form the UK. *SemanticScholar.org*. Retrieved from https://www.semanticscholar.org/paper/The-Effect-of-Indoor-Air-Pollution-on-Cognitive-the-Roth/6d210d6b55f2a3e908047ede9a716455c9ff34d4

Sanders NJ. (2012). What Doesn't Kill You Makes You Weaker: Prenatal Pollution Exposure and Educational Outcomes. *Journal of Human Resources, 47*(3), 826-850. Retrieved from https://jhr.uwpress.org/content/47/3/826

Sidney S, Quesenberry CP, Jaffe MH, Sorel M, Nguyen-Huynh M, Kushi L, Go A and Rana J. (2016). Recent Trends in Cardiovascular Mortality in the United States and Public Health Goals. *JAMA Cardiology, 1*(5), 594-599. doi:10.1001/jamacardio.2016.1326

Smith VK, Pattanayak SK and Van Houtven GL. (2006). Structural benefit transfer: an example using VSL estimates. *Ecological Economics, 60*, 361-371.

Smith VK, Van Houtven G and Pattanayak SK. (2002). Benefit Transfer via Preference Calibration: "Prudential Algebra" for Policy. *Land Economics, 78*(1), 132-152. doi:https://doi.org/10.2307/3146928

Soni A. (2008). *Allergic Rhinitis: Trends in use and expenditures, 2000 and 2005.* Agency for Healthcare Research and Quality, Medical Expenditure Panel Survey.

Stanley TD and Doucouliagos H. (2016). *Meta-Regression Analysis in Economics and Business.* New York, NY: Routledge. Retrieved from https://www.routledge.com/Meta-Regression-Analysis-in-Economics-and-Business/Stanley-Doucouliagos/p/book/9781138241145

Stephens M and Toohey D. (2022). The Impact of Health on Labor Market Outcomes: Evidence from a Large-Scale Health Experiment. *American Economic Journal: Applied Economics, 143*, 367-99.

Sturza J. (2010). A Review and Meta-Analysis of Utility Values for Lung Cancer. *Medical Decision Making, 30*(6), 685-693. doi:10.1177/0272989X10369004

Sun H, Shin YM, Xia M, KE S, Wan M, Yuan L, Guo Y and Archibald AT. (2022, November 9). Spatial Resolved Surface Ozone with Urban and Rural Differentiation during 1990-2019: A Space-Time Bayesian Neural Network Downscaler. *Environmental Science and Technology*, 7337-7349. Retrieved from https://pubs.acs.org/doi/10.1021/acs.est.1c04797

Thurston GD and Bell ML. (2021). The Human Health Co-Benefits of Air Quality Improvements Associated with Climate Change Mitigation. In E. Pinkerton KE and Rom WN, *Climate Change and Global Public Health* (pp. 181-202). Springer International Publishing.

Tipton E, Hartman E. (2023). Generalizability and trasnportability. In J. Zibizaretta, E. Stuart, D. Small, & P. Rosenbaum, *Handbook of Matching and Weighting Adjustments for Causal Inference.* Boca Raton, FL: CRC Press.

U.S. Bureau of Labor Statistics. (2015). *Current Population Survey.* 2015 Annual Social and. Retrieved from https://www.bls.gov/opub/reports/womensdatabook/2016/home.htm

U.S. Department of Health and Human Services. (2016). *Guidelines for Regulatory Impact Analysis.* Office of the Assistant Secretary for Planning and Evaluation . Retrieved from https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//171981/HHS_RIAGuidance.pdf

U.S. Department of Health and Human Services. (2017a). *Estimating Medical Costs for Regulatory Benefit-Cost Analysis: Conceptual Framework and Best Practices.* Prepared by Robinson LA, Rein DB and Hammitt JK. U.S. Health and Human Services. Retrieved from https://aspe.hhs.gov/reports/estimating-medical-costs-regulatory-benefit-cost-analysis-conceptual-framework-best-practices

U.S. Department of Health and Human Services. (2017b). *Valuing Time in U.S. Department of Health and Human Services Regulatory Impact Analyses: Conceptual Framework and Best Practices.* U.S. Department of Health and Human Services. Retrieved from https://aspe.hhs.gov/reports/valuing-time-us-department-health-human-services-regulatory-impact-analyses-conceptual-framework

U.S. Department of Health and Human Services. (2021). *Appendix D: Updating Value per Statistical Life (VSL) Estimates for Inflation and Changes in Real Income.* Retrieved from https://aspe.hhs.gov/reports/updating-vsl-estimates

U.S. Department of Transportation. (2021). *Departmental Guidance: Treatment of the Value of Preventing Fatalities and Injuries in Preparing Economic Analyses.* Retrieved from https://www.transportation.gov/sites/dot.gov/files/2021-03/DOT%20VSL%20Guidance%20-%202021%20Update.pdf

U.S. Department of Transportation. (2021a). *Departmental Guidance on Valuation of a Statistical Life in Economic Analysis*. Retrieved from Transportation Policy: https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-on-valuation-of-a-statistical-life-in-economic-analysis

U.S. Department of Transportation. (2021b). *Departmental Guidance*. Retrieved from Treatment of the Value of Preventing Fatalities and Injuries in Preparing Economic Analyses: https://www.transportation.gov/sites/dot.gov/files/2021-03/DOT%20VSL%20Guidance%20-%202021%20Update.pdf

U.S. Environmental Protection Agency. (1997). *The Benefits and Costs of the Clean Air Act, 1970-1990.* EPA. Retrieved from https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1970-1990-retrospective-study

U.S. Environmental Protection Agency. (1999). *The Benefits and Costs of the Clean Air Act, 1990 to 2010 - First Prospective Study.* EPA-410-R-99-001. Retrieved from https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2010-first-prospective-study

U.S. Environmental Protection Agency. (2000). *Regulatory Impact Analysis: Heavy-Duty Engine and Vehicle Standards and Highway Diesel Fuel Sulfur Control Requirements.* EPA Report 420-R-00-026. Retrieved from https://nepis.epa.gov/Exe/ZyNET.exe/P100K576.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2000+Thru+2005&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=

U.S. Environmental Protection Agency. (2009). *Integrated Science Assessment (ISA) for Particulate Matter.* EPA/600/R-08/139F. Retrieved from https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=216546

U.S. Environmental Protection Agency. (2011). *Benefits and Costs of the Clean Air Act 1990-2020, the Second Prospective Study.* U.S. EPA. Retrieved from https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2020-report-documents-and-graphics

U.S. Environmental Protection Agency. (2012). *Regulatory Impact Analysis for the Final Revisions to the National Ambient Air Quality Standards for Particulate Matter.* EPA-452/R-12-005 . Retrieved from https://www.epa.gov/sites/default/files/2020-07/documents/naaqs-pm_ria_final_2012-12.pdf

U.S. Environmental Protection Agency. (2014a). *Control of Air Pollution from Motor Vehicles: Tier 2 Motor Vehicle Emission and Fuel Standards Final Rule Regulatory Impact Analysis.* EPA Report 420-R-14-005. Retrieved from https://nepis.epa.gov/Exe/ZyPDF.cgi/P100ISWM.PDF?Dockey=P100ISWM.PDF

U.S. Environmental Protection Agency. (2014b). *Guidelines for Preparing Economic Analyses.* EPA National Center for Environmental Economics, Washington DC. Retrieved from https://www.epa.gov/sites/default/files/2017-08/documents/ee-0568-50.pdf

U.S. Environmental Protection Agency. (2019). *Integrated Science Assessment (ISA) for Particulate Matter.* EPA/600/R-19/188. Retrieved from https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=347534#tab-3

U.S. Environmental Protection Agency. (2020a). *Handbook on Valuing Changes in Time Use Induced by Regulatory Requirements and Other EPA Actions.* Washington DC: EPA National Center for Environmental Economics. Retrieved from https://www.epa.gov/sites/default/files/2020-12/documents/epa_handbook_on_valuing_changes_in_time_use_121520_final_508.pdf

U.S. Environmental Protection Agency. (2020b). *Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants.* Washington, DC: EPA/600/R-20/012. Retrieved from https://www.epa.gov/isa/integrated-science-assessment-isa-ozone-and-related-photochemical-oxidants

U.S. Environmental Protection Agency. (2022a). *Regulatory Impact Analysis for the Proposed Reconsideration of the National Ambient Air Quality Standards for Particulate Matter.* EPA Report: EPA-452/P-22-001. Retrieved from https://www.epa.gov/system/files/documents/2023-01/naaqs-pm_ria_proposed_2022-12.pdf

U.S. Environmental Protection Agency. (2022b). Supplement to the 2019 Integrated Science Assessment for Particulate Matter (Final Report, 2022). Retrieved from https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=354490

U.S. Environmental Protection Agency. (2022c). *Supplementary Material for the Regulatory Impact Analysis for "Standards of Performance for New, Reconstructed, and Modified Sources and Emissions Guidelines for Existing Sources: Oil and Natural Gas Sector Climate Review.* EPA-HQ-OAR-2021-0317. EPA External Review Draft of Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances.

U.S. Environmental Protection Agency. (2023). *Estimating PM2.5- and Ozone-Attributable Health Benefits.* Office of Air and Radiation. EPA-HQ-OAR-2019-0587. Retrieved from https://www.epa.gov/system/files/documents/2023-01/Estimating%20PM2.5-%20and%20Ozone-Attributable%20Health%20Benefits%20TSD_0.pdf

U.S. Environmental Protection Agency. (n.d.). *BenMAP - Benefits Mapping and Analysis Program website*. (Office of Air and Radiation) Retrieved from https://benmap.epa.gov/

U.S. Environmental Protection Agency. (December 2022). *Environmental Benefits and Analysis Mapping Program: User's Manual.* Retrieved from https://sab.epa.gov/ords/sab/f?p=114:0:32679603711722:APPLICATION_PROCESS=AGENCY_REVIEW:::AR_ID:2489

U.S. Health and Human Services. (2016). *Guidelines for Regulatory Impact Analysis.* Retrieved from https://aspe.hhs.gov/sites/default/files/private/pdf/242926/HHS_RIAGuidance.pdf

Valentine JC. (2019). Incorporating judgments about study quality into research syntheses. In H. Cooper, L. V. Hedges, & J. C. Valentine, *The Handbook of Research Synthesis and Meta-Analysis* (pp. 129-140). New York, NY: Russell Sage Foundation.

Viscusi WK. (2004). The value of life: estimates with risks by occupation and industry. *Economic Inquiry, 42*(1), 29-48.

Viscusi WK. (2015). The role of publication selection bias in estimates of the value of statistical life. *American Journal of Health Economics, 1*(1), 27-52.

Viscusi WK. (2018). Best estimate selection bias in the value of a statistical life. *Journal of Benefit-Cost Anaysis, 9*(2), 205-246.

Viscusi WK and Aldy JE. (2003). The value of a statistical life: a critical review of market estimates throughout the world. *Journal of Risk and Uncertainty, 27*, 5-76.

Viscusi WK and Masterman CJ. (2017). Income Elasticities and Global Values of a Statistical Life. *Journal of Benefit-Cost Analysis, 8*(2), 226-250.

Voorhees AS, Fann N, Fulcher C, Dolwick P, Hubbell B, Bierwagen B, Morefield P. (2011, February 15). Climate change-related temperature impacts on warm season heat mortality: a proof-of-concept methodology using BenMAP. *Environmental Science and Technology*, 1450-7. doi:10.1021/es102820y

Voorheis J. (2017). *Air Quality, Human Capital Formation and the Long-term Effects of Environmenal Inequality at Birth.* Census Working Papers, U.S. Census Bureau. Working Paper Number CARRA-WP-2017-05. Retrieved from https://www.census.gov/library/working-papers/2017/adrm/carra-wp-2017-05.html

Wang T and Sun F. (2022). Global gridded GDP data set consistent with the shared socioeconomic pathways. *Scientific Data, 9*, Article 221. doi:https://doi.org/10.1038/s41597-022-01300-x

Wang Y, Apte JS, Hill JD, Ivey CE, Patterson RF, Robinson AL, Tessum CW and Marshall JD. (2022). Location-specific strategies for eliminating US national racial-ethnic PM2.5 exposure inequality. *Proceedings of the National Academy of Sciences, 119*(44), e2205548119. doi:https://doi.org/10.1073/pnas.2205548119

Wei Y, Coull B, Koutrakis P, Yang J, Li L, Zanobetti A and Schwartz J . (2021). Assessing additive effect of air pollutants on morality rate in Massachusetts. *Environmental Health, 20*(1), 19.

Wilkinson MD, Dumontier M, Aalbersberg ID +50. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*(Article number: 160018). Retrieved from https://www.nature.com/articles/sdata201618

Williams AM and Phaneuf DJ. (2019). The Morbidity Costs of Air Pollution: Evidence from Spending on Chronic Respiratory Conditions. *Environmental and Resource Economics, 74*, 571-603. Retrieved from https://link.springer.com/article/10.1007/s10640-019-00336-9

Williams, P., & Hooten, M. (2016). Combining statistical inference and decisions in ecology. *Ecological Applications, 26*(6), 1930-1942.

Winer BJ. (1978). Statistics and data analysis: trading bias for reduced mean squared error. *Review of Psychology, 29*, 647-681.

Woods & Poole Economics. (2015). Woods & Poole Economics. *Proprietary Database*. Washington DC. Retrieved from https://www.woodsandpoole.com/

Yazdi MD, Wang Y, Di Q, Wei Y, Requia WJ, Shi L +6. (2021, Feb 22). Long-Term Association of Air Pollution and Hospital Admissions Among Medicare Participants Using a Doubly Robust Additive Model. *Circulation*. Retrieved from https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.120.050252

Young F, Capewell S, Ford E and Critchley J. (2010, September). Coronary mortality declines in the U.S. between 1980 and 2000 quantifying the contributions from primary and secondary prevention. *American Journal of Preventive Medicine, 39*(3), 228-234. doi:10.1016/j.amepre.2010.05.009

## Appendix A: Pooling Risk Estimates

The Agency asked for suggestions about how to pool estimates from studies with different underlying population characteristics. This is a major challenge and we do not have a complete solution. As a starting point for addressing this challenge, we offer the following example to illustrate how multiple estimates can be pooled while accounting for differences in the age ranges of the sampled individuals. Specifically, we use three estimates of PM2.5 mortality risks shown in Table 10 in the TSD and reproduced below:

| **Study information** | **Ages** | **Beta coefficient (se)** |
|---|---|---|
| Wu et al. (2000) | 65-99 | 0.0064 (0.0003) |
| Pope III et al. (2019) | 18-99 | 0.0133 (0.0016) |
| Turner et al. (2016) | 30-99 | 0.0058 (0.0010) |

We assume the risk coefficient in each study, $\beta$, is a reduced-form parameter in a Cox proportional hazards model that is a weighted average of latent parameters, $\theta$, for each age range covered by that study:

$$h = h_0 e^{-\beta x} = h_0 e^{-(\theta_1 A_1 + \theta_2 A_2 + \theta_3 A_3)x}, \quad \text{(A1)}$$

where $h_0$ is the background hazard rate when the pollutant concentration $x = 0$, and $A_1$, $A_2$, and $A_3$ are indicator variables for three mutually exclusive age ranges that cover the full range of ages considered in all three studies. Let $f_{i,j}$ be the fraction of the sample in study $i$ that falls into age category $j$. The table below shows the estimates from each study as weighted averages of three age ranges.

| **Study information** | **Equation** |
|---|---|
| Wu et al. (2000) | $0.0064 = f_{1,1}\theta_1 + f_{1,2}\theta_2 + f_{1,3}\theta_3$ |
| Pope III et al. (2019) | $0.0133 = f_{2,1}\theta_1 + f_{2,2}\theta_2 + f_{2,3}\theta_3$ |
| Turner et al. (2016) | $0.0058 = f_{3,1}\theta_1 + f_{3,2}\theta_2 + f_{3,3}\theta_3$ |

This forms a system of three equations, which can be written in matrix form and solved for $\theta$ as follows:

$$F\theta = B \Rightarrow \theta = F^{-1}B, \quad \text{(A2)}$$

where

$$F = \begin{bmatrix} f_{1,1} & f_{1,2} & f_{1,3} \\ f_{2,1} & f_{2,2} & f_{2,3} \\ f_{3,1} & f_{3,2} & f_{3,3} \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}, \text{ and } \quad B = \begin{bmatrix} 0.0064 \\ 0.0133 \\ 0.0058 \end{bmatrix}. \quad \text{(A3)}$$

To complete the example, we will assume that the age distribution in each study matched that of the U.S. general population (within the age range covered by each study), using the following age categories: $A_1 =$ 20-29 years old, $A_2 =$ 30-64 years old, and $A_3 =$ 65-99 years old. This gives:

$$F = \begin{bmatrix} 0.000 & 0.000 & 1.000 \\ 0.176 & 0.601 & 0.224 \\ 0.000 & 0.729 & 0.182 \end{bmatrix}$$

Using these assumptions, we get:

| Age range | Latent risk coefficient, $\theta$ | s.e. |
|---|---|---|
| 20-29 | 0.046 | 0.010 |
| 30-64 | 0.0064 | 0.0014 |
| 65-99 | 0.0064 | 0.00030 |

We estimated the standard errors using a Krinsky-Robb procedure, drawing from the sampling distributions of each coefficient estimate and re-solving the system of linear equations each time.

The results suggest that the risk for the youngest age range, 20-29 years old, is nearly an order of magnitude greater than that for older ages. This seems implausible in light of what is known about human physiology and aging, but we can see where it comes from mathematically: The Pope estimate for 18-99 year olds is 0.0133 while the Turner estimate for 30-99 year olds is 0.0058. The only way to rationalize these results is for the Pope estimate to be a weighted average of a much higher risk for 18-29 year olds and a lower risk for 30-99 year olds to produce the intermediate value of 0.0133.

This example highlights the strong assumptions that are required for the calibrated risk parameters to be consistent estimates of the true parameters when the pooled studies differ along more dimensions than those explicitly included in the pooling estimator. In this case, we would have to maintain either of the following assumptions:

1) age is the only relevant factor that differentiate the samples—or, in a linear model, perfectly uncorrelated with other unobserved factors that influence mortality risks—in which case the calibrated age coefficients are causal parameters and valid for predicting mortality risk changes, or
2) age is perfectly correlated with unobserved factors that differentiate the samples, in which case the age-specific latent risk coefficients are not causal parameters but still would provide accurate predictions of mortality risk changes.

In the foregoing example neither of these assumptions seems credible, so we do not recommend using the quantitative results of this example for policy analysis. To develop a more credible risk transfer function, multiple factors that differentiate the primary studies would need to be controlled in the pooling estimator. Our example provides only a starting point, not a complete solution to the challenge of pooling estimates from studies with different underlying population characteristics.

To generalize the illustrated approach, the analyst would begin by specifying a structural equation relating the estimated risk coefficients to all relevant and observable moderating factors that distinguish the primary studies, analogous to equation (A1) but including more factors than just age categories. Next, a suitable numerical method would be used to find the coefficients of the structural

equation—the $\theta$'s in equation (A1)—that minimize the sum of (possibly weighted) squared errors between the observed and predicted risk coefficients from the primary studies. To examine the robustness of the resulting coefficient estimates, sensitivity analysis could be used to quantify the potential bias associated with the full range of possible correlations between unobserved factors and observed factors in the structural meta-regression equation (Cinelli & Hazlett, 2020).

*R script*

```
set.seed(1234)
F <- matrix(c(0.000 , 0.000 , 1.000 ,
          0.176 , 0.601 , 0.224 ,
          0.000 , 0.729 , 0.182),3,3,byrow=TRUE)
B  <- c(0.0064 , 0.0133 , 0.0058)
SE <- c(0.0003 , 0.0016 , 0.0010)
theta <- solve(F)%*%B
# Krinsky-Robb:
MC <- 10000
theta.MC <- matrix(0,MC,3)
for(mc in 1:MC){
  Bmc <- rnorm(3,B,SE)
  theta.MC[mc,] <- solve(F)%*%Bmc }
```

## Appendix B: Validity Weighting in Meta-Analysis

A common challenge when pooling evidence involves the assignment of weights on individual estimates from different studies. Conventional methods of meta-analysis pool estimates using weights that depend on the estimated sampling variability of each primary estimate, typically reported as standard errors, and the overall heterogeneity among the estimates (Borenstein M, Hedges LV, Higgins JPT, Rothstein H, 2010). However, beyond sampling variability and heterogeneity, primary studies considered for a meta-analysis may also differ on other important dimensions, including the validity of their identification strategies or other relevant design features that may influence their credibility. A more efficient study with low standard errors might be based on an identification strategy that is less likely to isolate a causal treatment effect that is generalizable to the broader population or transportable to other settings (Tipton E, Hartman E, 2023) than another study with higher standard errors, which would set up a trade-off between precision and accuracy when combining the evidence from the two studies. Typically, differences in study validity are considered only in the data collection phase when analysts decide which studies to include or exclude. As noted by Cordray and Morphy (2009), "Agreement on the exclusion of studies, however, is not widespread." Many studies have strengths and weaknesses that make their overall credibility difficult to judge. In these cases, which may comprise a large share of the relevant studies, the include-or-exclude determination may itself be uncertain. In light of this common dilemma, it would be advantageous to extend standard meta-analytic methods to relax the strict in-or-out constraint to handle borderline cases.

One way to do this would be to assign a "validity weight" to each relevant study representing the analysts' considered judgments regarding the strength of the identification strategy and other elements of the study design.[15] If the study design is known to be valid in all relevant respects and thereby provides an unbiased estimate of the treatment effect that can be applied to the policy context of interest, then the study would be assigned a validity weight of 1. If the study is known to be invalid on one or more relevant dimensions and therefore biased by some unknown amount, the study would be assigned a validity weight of 0. In these extreme cases, the include-or-exclude decision is clear. When the analysts cannot make a definitive 0-1 validity determination, they would assign an intermediate weight between 0 and 1 representing their subjective judgments about the overall validity of the study. These judgments would be made prior to analyzing the meta-data and would be treated as exogenous features of the studies in the analysis stage. Then, to combine the evidence from the full set of relevant studies, optimal weights that balance the differential precision and accuracy of the studies can be derived that include not only the point estimates and standard errors from each study (the usual ingredients of any fixed- or random-effects meta-analysis) but also the assigned validity weights.

[15] Valentine (2019) discussed incorporating judgments about study quality in meta-analysis and recommended using multi-dimensional measures of study quality as moderators in a meta-regression rather than a single dimensional measure of quality to derive weights like the approach we propose here. Valentine's recommendation was partly informed by Ahn and Becker (2011), who conducted a simulation study using a quality weighted estimator of the form $\frac{\sum q_i w_i d_i}{\sum q_i w_i}$, where $d_i$ is the effect size, $w_i$ is the conventional inverse variance weight, and $q_i$ is the subjective quality weight for study $i$. Ahn and Baker found that this estimator did not eliminate bias but had a lower MSE than the conventional fixed-effects estimator when lower quality studies had larger biases, but they nevertheless concluded that "quality weights lead to bias in almost every condition studied, [so] we recommend against the use of quality weights" (p 555).

*Model set-up*

We will denote the target of estimation as $Y$, which could be a hazard ratio, dose-response coefficient, or some other measure of the causal effect of a specific air pollutant on a specific human health outcome. We assume that each of $i = 1,2, \ldots, N$ studies reported a single estimate of $Y$, $y_i$, with a standard error, $se_i$. The analysts assign a subjective "validity weight," $\rho_i$, to each study representing their judgment about the validity of the identification strategy used in study $i$: if a study is credible then $\mathbb{E}[y_i] = Y$, otherwise $E[y_i] = Z$ where $Z \neq Y$. I.e., poorly designed studies—those that do not use estimators with sampling distributions centered on the true causal effect—are biased by some unknown amount $Z - Y$.[16]

*The estimator*

Consider the family of convex estimators:

$$\hat{Y} = \sum w_i y_i \quad \text{where} \quad \sum w_i = 1 \quad \text{and} \quad 0 \leq w_i \leq 1 \ \ \forall i. \tag{B1}$$

We want to find the set of weights that gives the lowest mean squared error for $\hat{Y}$.[17] In expectation, we have

$$\mathbb{E}\left[\hat{Y}\right] = \mathbb{E}\left[\sum w_i y_i\right] = \sum w_i[\rho_i Y + (1 - \rho_i)Z]. \tag{B2}$$

The mean squared error of the estimator is

$$MSE = \mathbb{E}\left\{\sum w_i[\rho_i Y + (1 - \rho_i)Z] - Y\right\}^2 + \sum w_i^2 se_i^2. \tag{B3}$$

Including the constraint that the weights must sum to one, the Lagrangian is

$$\mathcal{L} = \mathbb{E}\left\{\sum w_i[\rho_i Y + (1 - \rho_i)Z] - Y\right\}^2 + \sum w_i^2 se_i^2 + \lambda\left(\sum w_i - 1\right). \tag{B4}$$

After combining the first-order conditions for an optimum of the Lagrangian with the constraint to eliminate the Lagrange multiplier, we arrive at the following formula for the optimal weights:

$$w_i = \frac{se_i^{-2}}{\sum se_i^{-2}} + A(w)\left\{\frac{se_i^{-2}}{\sum se_i^{-2}}\left(\sum \frac{[\rho_i Y + (1 - \rho_i)Z]}{se_i^2}\right) - \frac{[\rho_i Y + (1 - \rho_i)Z]}{se_i^2}\right\}, \tag{B5}$$

[16] A more general set-up would also allow for (zero-mean) non-sampling errors as in a conventional random-effects meta-analysis, and would allow the variances of non-sampling errors to differ across credible and non-credible studies.

[17] The mean-squared-error (MSE) metric combines into a single measure two of the most important criteria for estimators: bias and variance (Kennedy P, 2008). Rather than focusing on the least variable estimators among unbiased or consistent estimators, the MSE criterion allows a smooth trade-off between the two. One implication of using the MSE criterion is that the researcher allows the use of an inconsistent estimator even when a consistent estimator may be available (Winer BJ, 1978). This can be an advantage if it is possible to achieve a large reduction in variance by accepting a small increase in bias. In these cases, the researcher can be "less wrong" on average. We can make this idea more concrete by considering the estimation problem as one of minimizing a loss function (Williams & Hooten, 2016): what is the cost of our point estimate deviating from the true value? In the present context, this cost might come in the form of deadweight loss from a poorly designed environmental regulation owing to the use of an erroneous point estimate of an important epidemiological or economic parameter. A system or market that the decision-maker is trying to regulate does not know or care if the estimation error is due to bias or variance. The MSE loss function effectively assumes that the cost of the estimation error increases with the square of the difference between the point estimate used to design the policy and the true value of the parameter, which can be viewed as a second-order approximation of any loss function that has a minimum at the true parameter value.

where

$$A(w) = \sum w_i[\rho_i Y + (1 - \rho_i)Z] - Y. \quad \text{(B6)}$$

The first term on the right-hand side of equation (B5) is the familiar inverse-variance precision weight from a conventional fixed-effect meta-analysis. The second term is the adjustment to the weights for the differential validity assignments across the studies. It is not possible compute the $w_i$'s straight away using equation (B5) since the $w_i$'s also appear on the right-hand side, embedded in $A(w)$. But it is straightforward to iterate equation (B5) to a fixed-point, replacing negative weights with zeros as needed.[18]

To compute the weights, we first need estimates of $Y$ and $Z$. To estimate these quantities, consider

$$\mathbb{E}[y_i] = \rho_i Y + (1 - \rho_i)Z = Z + \rho_i(Y - Z). \quad \text{(B7)}$$

Therefore, to compute initial estimates of $Y$ and $Z$, we can regress $y_i$ on a constant and $\rho_i$ and then take the intercept as an initial estimate of $Z$ and take the slope as an initial estimate of $Y - Z$. We can then use these initial estimates to compute $w_i$'s, then use the $w_i$'s to compute $\hat{Y}$, which is an updated estimate of $Y$. Then we can compute an updated estimate of $Z$ using

$$\hat{Z} = \frac{\sum y_i - \hat{Y} \sum \rho_i}{\sum (1 - \rho_i)} \quad \text{(B8)}$$

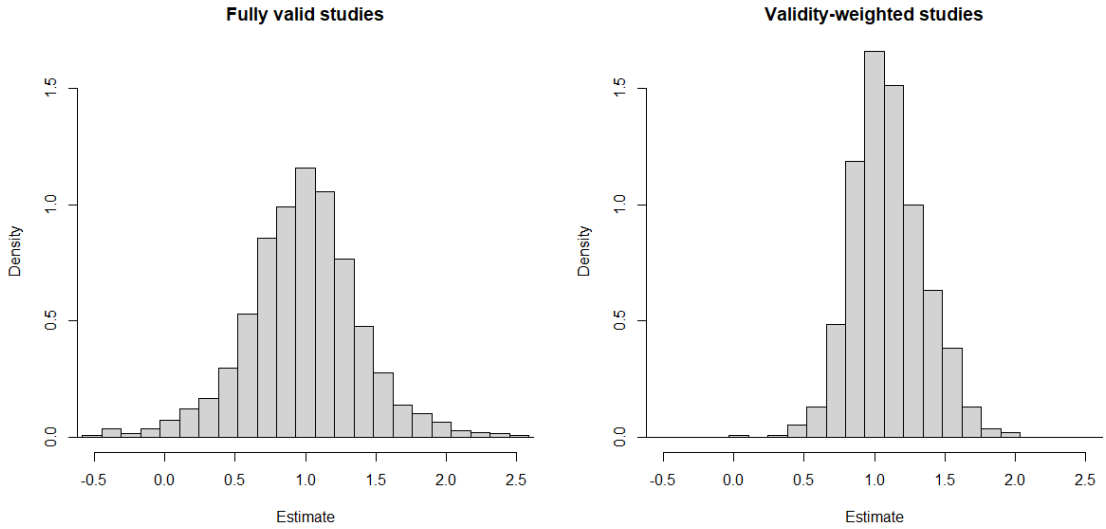
and repeat until convergence.

*Example using constructed data*

In the R script below, we compare the performance of the proposed validity-weighted estimator to the conventional precision-weighted meta-analysis estimator that uses only fully valid studies (i.e., studies with $\rho_i = 1$) and discards the rest. In this example, $Y = 1, Z = 2$, there are 10 studies with standard errors between 0.2 and 1.0, and the credibility weights are $\rho = [1, 1, .9, .9, .8, .8, .7, .7, .6, .6]$. That is, the analyst has collected two fully valid studies, two less valid studies, and so on, down to two marginally valid studies. We simulated data using these assumptions 500 times and computed both estimators for each simulated dataset.

The histograms below show the sampling distributions for the conventional meta-estimator using only fully valid studies (left) and the validity-weighted meta-estimator (right). The MSEs for the conventional and validity-weighted estimators are 0.161 and 0.073, respectively. On close inspection, the bias-variance tradeoff is apparent here. The validity-weighted estimator is slightly biased away from $Y = 1$ towards $Z = 2$ (its mean is 1.095 compared to 0.982 for the conventional estimator), but it has a substantially lower standard error (0.253 compared to 0.401).

[18] As a cross-check, we also used a numerical gradient algorithm to compute the weights while respecting all of the constraints, independent of the formulas given in equations (B5) and (B6). In all of the examples we tried, both methods gave virtually identical weights.




The validity-weighted estimator also will indicate which of the candidate studies should be excluded from the meta-analysis by producing weights for those studies equal to zero. In the example above, on average 1.43 of the 10 studies were excluded by the validity-weighted estimator. There also appears to be an intermediate level of validity below which studies should be excluded ex ante. When we ran the example again excluding the two $\rho_i = 0.6$ studies the MSE went down, then when we also excluded the two $\rho_i = 0.7$ studies the MSE went back up. So in this case the cutoff appears to be between 0.6 and 0.7. We suspect this result is not fully general, but we would conjecture that some such interior value will be optimal in many relevant cases.

*R script*

```
#-------------------------------------------------------------------------------
# FUNCTIONS:
#-------------------------------------------------------------------------------
{

 pause  <- function(sec){

  if(missing(sec)){
   readline(prompt='Paused. Press [enter] to continue.\n')
  }else{
   Sys.sleep(sec)
  }

 }

 MSE.fn <- function(w,Y,Z,se,rho){
  MSE <- (sum(w*(rho*Y+(1-rho)*Z))-Y)^2 + sum(w^2*se^2)
  return(MSE)
 }

 w.fn   <- function(Y,Z,se,rho){
```

```
if(method=='comp'){

  N    <- length(se)
  w    <- matrix(1/N,N,1)
  MSE  <- MSE.fn(w,Y,Z,se,rho)
  done <- 0
  while(done == 0){

    # up
    MSE.up <- MSE
    w.up   <- w
    for(i in 1:N){
      w.ij <- w
      if(w.ij[i]<=(1-dw)){
        w.ij[i] <- w.ij[i] + dw
        for(j in 1:N){
          if(j!=i & w.ij[j]>=(dw+dw/10)){
            w.ij[j] <- w.ij[j] - dw
            MSE.ij  <- MSE.fn(w.ij,Y,Z,se,rho)
            if(MSE.ij < MSE.up){w.up <- w.ij; MSE.up <- MSE.ij}
            w.ij[j] <- w.ij[j] + dw
          }
        }
        w.ij[i] <- w.ij[i] - dw
      }
    }

    # down
    MSE.dn <- MSE.up
    w.dn   <- w.up
    for(i in 1:N){
      w.ij <- w
      if(w.ij[i]>=(dw+dw/10)){
        w.ij[i] <- w.ij[i] - dw
        for(j in 1:N){
          if(j!=i & w.ij[j]<=(1-dw)){
            w.ij[j] <- w.ij[j] + dw
            MSE.ij  <- MSE.fn(w.ij,Y,Z,se,rho)
            if(MSE.ij < MSE.dn){w.dn <- w.ij; MSE.dn <- MSE.ij}
            w.ij[j] <- w.ij[j] - dw
          }
        }
        w.ij[i] <- w.ij[i] + dw
      }
    }
```

```
      if(MSE.dn < MSE){MSE <- MSE.dn; w <- w.dn}else{done <- 1}

    }

  }

  if(method=='math'){

    phi <- .25

    N <- length(se)
    w <- matrix(1/N,N,1)

    done <- 0
    while(done==0){

      Aw <- sum(w*(rho*Y+(1-rho)*Z))-Y
      ww <- se^(-2)/sum(se^(-2)) + Aw*( se^(-2)/sum(se^(-2))*sum((rho*Y+(1-rho)*Z)/se^2) - (rho*Y+(1-
rho)*Z)/se^2 )

      ww[which(ww<0)] <- 0
      ww[which(ww>1)] <- 1
      ww <- ww/sum(ww)

      # print(ww); pause()

      if(max(abs((w-ww)/(w+1e-12)))<0.0001){done <- 1}else{w <- (1-phi)*w + phi*ww}

    }

  }

  return(w)

 }

}


#---------------------------------------------------------------------------
# MAIN PROGRAM:
#---------------------------------------------------------------------------

set.seed(1234)
```

```
Y    <- 1
Z    <- 2
se.lo <- 0.2
se.hi <- 1.0
rho   <- c(1,1,.9,.9,.8,.8,.7,.7,.6,.6) # MSE = 0.079
# rho   <- c(1,1,.9,.9,.8,.8,.7,.7)      # MSE = 0.059
# rho   <- c(1,1,.9,.9,.8,.8)            # MSE = 0.074
N    <- length(rho)

method <- 'math' # 'comp' or 'math

dw   <- .005

MC      <- 1000
Y.infMC  <- matrix(0,MC,1)
Y.hatMC  <- matrix(0,MC,1)
YY.hatMC <- matrix(0,MC,1)
n.MC    <- matrix(0,MC,1)
for(mc in 1:MC){

 se <- runif(N,se.lo,se.hi)
 w  <- w.fn(Y,Z,se,rho) # ideal weights (infeasible)

 # Simulate data:
 y <- matrix(0,N,1)
 c <- matrix(0,N,1)
 for(i in 1:N){
  c[i] <- 1*(runif(1)<rho[i])
  y[i] <- c[i]*Y + (1-c[i])*Z + rnorm(1,0,se[i])
 }

 Y.infMC[mc] <- sum(w*y)

 w.hat0 <- matrix(1/N,N,1)
 x     <- cbind(matrix(1,N,1),matrix(rho,N,1))
 b     <- solve(t(x)%*%x)%*%t(x)%*%y
 Z.hat0 <- b[1]
 Y.hat0 <- b[2] + Z.hat0

 w.hat1 <- w.fn(Y.hat0,Z.hat0,se,rho)
 Y.hat1 <- sum(w.hat1*y)
 Z.hat1 <- (sum(y)-Y.hat1*sum(rho))/sum(1-rho)

 w.hat2 <- w.fn(Y.hat1,Z.hat1,se,rho)
 Y.hat2 <- sum(w.hat2*y)
 Z.hat2 <- (sum(y)-Y.hat2*sum(rho))/sum(1-rho)
```

```
  w.hat3 <- w.fn(Y.hat2,Z.hat2,se,rho)
  Y.hat3 <- sum(w.hat3*y)
  Z.hat3 <- (sum(y)-Y.hat3*sum(rho))/sum(1-rho)

  Y.hatMC[mc] <- Y.hat3
  n.MC[mc]    <- sum(w.hat3<=(dw+dw/10))

  yy <- y[which(rho==1)]
  ss <- se[which(rho==1)]

  ww.hat <- (1/ss^2)/sum((1/ss^2))

  YY.hat <- sum(ww.hat*yy)

  YY.hatMC[mc] <- YY.hat

  if(mc>5){
    par(mfrow=c(1,2))
    hist(YY.hatMC[1:mc],breaks=seq(-1,3,length=30),freq=FALSE,xlim=c(-.5,2.5),ylim=c(0,1.6),main='Fully
valid studies',xlab='Estimate')
    hist(Y.hatMC[1:mc], breaks=seq(-1,3,length=30),freq=FALSE,xlim=c(-
.5,2.5),ylim=c(0,1.6),main='Validity-weighted studies',xlab='Estimate')

    cat('\014Working on mc =',sprintf('%-.0f',mc),'of',sprintf('%-.0f\n',MC))
    cat('MSE[Yinf]  =',sprintf('%8.5f\n',mean((Y.infMC[1:mc] -Y)^2)))
    cat('MSE[Yhat]  =',sprintf('%8.5f\n',mean((Y.hatMC[1:mc] -Y)^2)))
    cat('MSE[YYhat] =',sprintf('%8.5f\n',mean((YY.hatMC[1:mc]-Y)^2)))
    cat('E[n]       =',sprintf('%8.5f\n',mean(n.MC[1:mc])))
    pause(.05)
  }else{
    cat('\014Working on mc =',sprintf('%-.0f',mc),'of',sprintf('%-.0f\n',MC))
  }

}
```