Roger O. McClellan, DVM, MMS, DSc (Honorary),
Diplomate-ABT, Diplomate-ABVT, Fellow-ATS
Advisor, Toxicology and Human Health Risk Analysis
13701 Quaking Aspen Place N.E.
Albuquerque, NM 87111-7168
Tel: 505-296-7083
E-mail: roger.o.mcclellan@att.net

August 18, 2009

Mr. Dennis Boal
Chairman, Environmental Quality Council
State of Wyoming
122 West 25th Street
Henschler Building , Room 1714
Cheyenne, WY 82002

Dear Mr. Boal:

I am pleased to have the opportunity to offer comments on the proposed petition for Air Quality Rulemaking entitled "Petition to Establish Primary and Secondary Wyoming Ambient Air Quality Standards for Ozone that are More Stringent than the National Ambient Air Quality Standards" submitted by Elaine Crumpley and others. The comments I offer represent my personal professional opinions on these matters drawing on my 4 decades of experience as a scientist engaged in conducting research on air quality issues and, most importantly, 3 decades of experience as an advisor to the U.S. Environmental Protection Agency (U.S. EPA) and other organizations on setting air quality standards and achieving attainment with those standards. I am currently a consultant to Shell Exploration and Production Company on air quality issues related to their operations on the Pinedale Anticline. However, I wish to emphasize the comments I offer on this matter are my own professional opinions and are not necessarily the views of Shell.

So you and the other Council members will be aware of my background and professional expertise, I am attaching a copy of my biography. My entire scientific career has been devoted to developing scientific information that will aid in providing guidance, including the setting of standards, that will help ensure the health and well-being of people while allowing them the benefits associated with a strong economy. I had the good fortune to be asked to advise the U.S. EPA on scientific matters related to air quality soon after the Agency was created.

One of my first advisory roles was to serve as the Chair of an Ad Hoc Committee to review the scientific basis of the first National Ambient Air Quality Standard for airborne lead. The U.S. EPA was required to set that standard under the authority granted it by Congress under

the Clean Air Act which was first enacted in 1970. In many ways the activities of the Lead Standard Review Committee served as a template for subsequent establishment of the Clean Air Act Scientific Advisory Committee (CASAC) by the Clean Air Act Amendments of 1977. Over the past 3 decades, I have served on numerous Panels of the CASAC advising the U.S. EPA Administrator on the science that informs the policy judgments required to establish standards for Ozone and the other criteria pollutants. I chaired the CASAC from 1988 to 1992. As a result of my advisory activities I have learned a lot about the Clean Air Act and how it is administered.

Let me briefly review how the Clean Air Act came into being and describe some key features. In the mid-1900s, increased concern developed for deteriorating air quality in the United States, especially in heavily industrialized areas and in major cities with increasing vehicle traffic. This concern stimulated passage of legislation by local and state governments to curb air pollution. This was followed by weak federal legislation. The result by the late 1960s was a hodge-podge of legislation and regulations, that in the view of many citizens, was not effective in improving air quality. It soon became apparent that if true progress were to be made in improving air quality across the United States, it was going to require a coordinated national effort that would also engage the states and local governments. The result was passage of the Clean Air Act (CAA) of 1970.

An important component of the CAA was the creation of an orderly system by which the U.S. EPA establishes National Ambient Air Quality Standards (NAAQS) for criteria pollutants like ozone. The criteria pollutants are pollutants that arise from multiple sources and are found across the United States. The clear intention of the CAA was to have a single National Standard for each of these criteria pollutants and, thus, avoid the confusion of multiple standards, i.e. potentially one standard in Wyoming and another in New York and yet another in Ohio. I will discuss later how these NAAQS are set.

A second key component of the CAA delegates authority to the individual states to conduct monitoring programs and develop State Implementation Plans. This requires each state to develop its own customized plan for attaining the NAAQS if it is determined that areas in the state do not meet the NAAQS. The CAA recognized that the individual states, with their knowledge of local conditions, would be in the best position to create strategic plans for attaining compliance with the NAAQS in their State.

As I noted earlier, I became involved soon after the CAA was passed in advisory roles in the setting of the NAAQS for the various criteria pollutants including ozone. After the initial

NAAQS were set in the early 1970s, the CAA mandated that they were to be reviewed every 5 years, a schedule that has rarely been achieved. The review of each NAAQS is an extraordinarily complex and resource intensive process. Any given review may take hundreds of person-years of effort by U.S. EPA staff and consultants over a period of 5 years or more.

The initial step of each NAAQS review is the compilation of everything that is known about the pollutant; what are the sources, how is it transported in the atmosphere, what are typical human exposures and what is the evidence for the pollutant at a given level of exposure causing health effects in human populations. All of this information is compiled by the EPA staff with the assistance of dozens of consultants into what are called Criteria Documents. These are frequently a thousand pages or more in length.

In a second step, the EPA staff, again with the assistance of many consultants, prepares a Staff Paper that documents how the information in the Criteria Document can be used in setting or revising the particular NAAQS. This includes considering alternatives for each of the four elements of a NAAQS; (a) the indicator, such as ozone for photochemical oxidants, (b) an averaging time, such as 8 hours or annual, (c) the specific level, such as 75 ppb, and (d) the statistical form, e.g. the standard is attained if the annual fourth-highest daily maximum 8-hour average concentration, averaged over 3 years is not exceeded.

The Criteria Documents and Staff Papers, sometimes several drafts, are reviewed by the EPA's Clean Air Scientific Advisory Committee (CASAC). Frequently, the CASAC, which consists of 7 members appointed by the EPA Administrator, is augmented with another dozen or more consultants. The CASAC then offers its advice to the EPA Administrator on how the science in the Criteria Document and Staff Paper should inform the Administrator's policy judgment in setting the four elements of the NAAQS.

At the next step, it is the responsibility of the EPA Administrator to issue a proposed rule for public comment. The proposed rule published in the Federal Register typically requests comments on a range of options. For example, the last proposed ozone rule published in the July 11, 2007 Federal Register solicited comments on "alternative levels down to 0.060 ppm and up to and including retaining the current 8-h standard of 0.08 ppm (effectively 0.084 using current data rounding conventions)." It is my understanding that the EPA received thousands of comments on the proposed rule, some suggested retaining the then current standard and others suggesting a much tighter standard.

I submitted my personal comments on the proposed rule and also joined with 8 scientific colleagues in submitting a report -" Critical Considerations in Evaluating Scientific Evidence of Health Effects of Ambient Ozone: A Conference Report" to the EPA Ozone Rule Docket (attached). That Report reviewed key scientific issues that must be considered in setting the ozone NAAQS. In addition, the Report and I, in my personal comments, emphasized --"that there is no scientific methodology that, in the absence of judgment, can define the precise numerical level, related averaging time, and statistical form of the NAAQS. The selection of these elements of the NAAQS involves policy judgments that should be informed by scientific information and analyses." Thus, neither I nor the Panel offered a policy judgment on the specific numerical level of the revised ozone NAAQS. In short, we did not think it appropriate for us, as scientist citizens, to take on the policy judgment role the CAA clearly delegates to the U.S. EPA Administrator.

The CAA very specifically assigns responsibility for the setting of NAAQS to the Administrator of the U.S. EPA. With regard to the setting of primary or health-based standards, it defines the standard as one "the attainment and maintenance of which, in the judgment of the Administrator, based on such criteria and allowing an adequate margin of safety, are requisite to protect public health." Thus, the responsibility for setting a NAAQS clearly rests with the EPA Administrator and involves judgment. The CAA, as amended, calls for establishing a Clean Air Scientific Advisory Committee (CASAC) to <u>advise</u> the Administrator on the science informing the Administrator's policy judgments. It is noteworthy that the CAA very wisely did not establish a Clean Air <u>Standard Setting</u> Committee.

A landmark decision rendered by the Supreme Court in "Whitman versus American Trucking Association" found that the U.S. EPA could not consider cost in setting the NAAQS. In that case, Justice Breyer opined that "this interpretation of section 109 does not require EPA to eliminate every health risk, however slight, at any economic cost, however great, to the point of "hurtling" industry over "the brink of ruin," or even, forcing "deindustrialization."

Breyer explained:

"The statute, by its express terms, does not compel the elimination of all risk; and it grants the Administrator sufficient flexibility to avoid setting ambient air quality standards ruinous to industry.

Section 109(b)(1) directs the Administrator to set standards that are "requisite to protect the public health" with "an adequate margin of safety." But these words do not describe a

world that is free of all risk – an impossible and undesirable objective (citation omitted). Nor are the words "requisite" and "public health" to be understood independent of context. We consider football equipment "safe" even if its use entails a level of risk that would make drinking water "unsafe" for consumption. And what counts as "requisite" to protecting the public health will similarly vary with background circumstances, such as the public's ordinary tolerance of the particular health risk in the particular context at issue. The Administrator can consider such background circumstances when "deciding what risks are acceptable in the world in which we live" (citation omitted).

The statute also permits the Administrator to take account of comparative health risks. That is to say, she may consider whether a proposed rule promotes safety overall. A rule likely to cause more harm to health than it prevents is not a rule that is "requisite to protect the public health." For example, as the Court of Appeals held and the parties do not contest, the Administrator has the authority to determine to what extent possible health risks stemming from reductions in tropospheric ozone (which, it is claimed, helps prevent cataracts and skin cancer) should be taken into account in setting the ambient air quality standard for ozone (citation omitted).

The statute ultimately specifies that the standard set must be "requisite to protect the public health" "in the judgment of the Administrator," § 109(b)(1), 84 Stat. 1680 (emphasis added), a phrase that grants the Administrator considerable discretionary standard-setting authority.

The statute's words, then, authorize the Administrator to consider the severity of a pollutant's potential adverse health effects, the number of those likely to be affected, the distribution of the adverse effects, and the uncertainties surrounding each estimate. (citation omitted). They permit the Administrator to take account of comparative health consequences. They allow her to take account of context when determining the acceptability of small risks to health. And they give her considerable discretion when she does so.

This discretion would seem sufficient to avoid the extreme results that some of the industry parties fear. After all, the EPA, in setting standards that "protect the public health" with "an adequate margin of safety," retains discretionary authority to avoid regulating risks that it reasonably concludes are trivial in context. Nor need regulation lead to deindustrialization. Pre-industrial society was not a very healthy society; hence a standard demanding the return of the Stone Age would not prove "requisite to protect the public health."

Although I rely more heavily than does the Court upon legislative history and alternative sources of statutory flexibility, I reach the same ultimate conclusion. Section 109 does not delegate to the EPA authority to base the national ambient air quality standards, in whole or in part, upon the economic costs of compliance."

Let me now turn specifically to the Petition of Elaine Crumpley and others. I have no doubt as to the good interactions of the petitioners. However, in my opinion, the Petitioners fail to appreciate the historical and statutory basis of the Clean Air Act and the specific roles assigned to the U.S. EPA and to the States.

The Clean Air Act as passed in 1970 was intended to replace a hodge-podge collection of local, state and federal laws and regulations that were proving ineffective in improving air quality. Since 1970, air quality across the United States has been improving. A cornerstone of the process is the establishment of National Ambient Air Quality Standards. As I have noted, that is a very resource intensive process. It is beyond my imagination as to how a State like Wyoming could propose to devote sufficient resources to the development of WAAQS for ozone and other pollutants that would provide a more defensible standard than the NAAQS.

In their Petition, Elaine Crumpley and others propose a very specific numerical standard, "at a level of 0.065 parts to per million, daily maximum 8-hour average." In proposing by Petition to set the WAAQS at 0.065 ppm ozone the Petitioners rely on the advisory opinion of the EPA's Clean Air Scientific Advisory Committee that a primary standard be set in the range of 0.060 to 0.070 ppm ozone. As I noted earlier, CASAC is an advisory committee with responsibility to provide advice to inform the EPA Administrator's policy judgments in setting the NAAQS. In my opinion, CASAC overstepped its statutory responsibility when it demands that the Administrator make a policy judgment that sets the NAAQS no higher than 0.070 ppm.

In petitioning for the setting of a WAAQS at 0.065 ppm ozone, the midpoint of the CASAC advisory range, the Petitioners have introduced their own policy judgment in the process. They offer no rationale for why they did not select the upper end of the range, 0.070 ppm ozone, or alternatively, the lower end of the range, 0.060 ppm ozone. The key point to be recognized is that science should inform the setting of NAAQS, however, science alone cannot define the precise numerical level and associated statistical form. If scientists, such as CASAC, advance their opinion on the science and then prescribe a specific numerical level within a narrow range the scientists are mixing the science with their personal "policy judgments" as to the level of the standard.

There was sound basis for the Clean Air Act not establishing a "Scientific Standard Setting Committee," but rather recognizing that policy judgments were involved in the setting of a NAAQS. In my opinion, it would not be wise for the State of Wyoming to use the policy judgments of the Petitioners, albeit based on the policy judgments of CASAC, in setting a WAAQS.

If the State of Wyoming were to develop a scheme for establishing WAAQS for ozone and other pollutants, it would be necessary for some individual, the Director of Wyoming DEQ, or some authorized body, such as EQC, to assume responsibility for making "policy judgments" in a manner analogous to that of the EPA Administrator. If that approach were taken, that duly authorized individual or individuals serving the State of Wyoming could make a "policy judgment" setting a WAAQS that was higher than, equal to, or lower than the NAAQS. Indeed, I think it is a reach on the part of the Petitioners to automatically assume that the WAAQS should be set at 0.065 ppm ozone in the absence of any administrative review involving "policy judgment" by a dispassionate administrator or administrative body whose concern is the welfare of the State of Wyoming and its citizens. In that regard, it is useful to recall the advice of Supreme Court Justice Breyer on setting standards.

It is apparent that if the Petition were accepted, it is conceivable that a WAAQS could be set at a level different from the NAAQS. This would immediately create tension between the dueling standards, the NAAQS and the WAAQS. The Clean Air Act and the NAAQS for ozone are the "law of the land." Hence, it is likely that the Wyoming DEQ will still need to proceed with development of a "State Implementation Plan" (SIP) for achieving state-wide attainment of the NAAQS for ozone by March 2013. The effort required to develop a defensible and effective SIP will be very substantial. If a WAAQS were established at a lower level than the NAAQS, it would be necessary to carry out a parallel effort to develop a strategic plan for attaining state-wide WAAQS. It cannot be assumed that the plan for attaining compliance with a lower WAAQS would only involve further reductions in emissions of ozone precursors from sources identified in the SIP to achieve attainment with the NAAQS. It is very likely that additional sources of emissions of ozone precursors would need to be identified. As the ozone standard moves closer to natural background levels, it will be even more difficult to achieve.

The U.S. EPA has already initiated a new review of the NAAQS for ozone with a target date of March 2013 for either reaffirming the existing NAAQS or issuing a new NAAQS. That process will consider again the scientific evidence available during the previous ozone NAAQS

review as well as newly published studies. It would be inappropriate to speculate on the outcome of that process which will again involve using "science" to inform a "policy judgment."

Let me now turn to my personal observations on efforts in Wyoming to improve air quality. I have been impressed by the quality and efforts of the personnel associated with the Wyoming DEQ and the oil and gas producers in Wyoming. Moreover, I have been impressed with the good faith efforts from these parties and local communities to make progress in reducing the emissions of ozone precursors and improve air quality. The effort and the progress are especially remarkable in view of the serious personnel resource limitations faced by the Wyoming DEQ.

In my view, activities to improve air quality across the State of Wyoming, and specifically in Sublette and Sweetwater Counties are on the right trajectory. Hence, I think the wise course of action is to "stay the course." In my opinion, I think acceptance of the Petition offered by Elaine Crumpley and others would have a high probability of having a negative impact on improving air quality across Wyoming and, moreover, could have other unanticipated negative impacts on the State of Wyoming as it struggles to meet the substantial demands associated with attaining the NAAQS and, potentially, a WAAQS if the Petition were accepted.

I am led to recommend that the Petition of Elaine Crumpley and others should be denied by the Environmental Quality Council.

Respectfully submitted,

Roger O. McClellan

Attachments:
Roger O. McClellan Biography
Preprint: Critical Considerations in Evaluating Scientific Evidence of Health Effects of Ambient Ozone: A Conference Report

BIOGRAPHY

ROGER O. McCLELLAN, DVM, MMS, DSc (Honorary),
Dipl-ABT, Dipl-ABVT, Fellow-ATS
Advisor: Inhalation Toxicology and Human Health Risk Analysis
13701 Quaking Aspen NE
Albuquerque, NM 87111-7168, USA
Tel: (505) 296-7083
Fax: (505) 296-9573
e-mail: roger.o.mcclellan@att.net

ROGER O. McCLELLAN is currently an advisor to public and private organizations on issues concerned with inhalation toxicology and human health risk analysis and safety evaluation of consumer products and pharmaceuticals. He received his Doctor of Veterinary Medicine degree with Highest Honors from Washington State University in 1960 and a Master of Management Science degree from the University of New Mexico in 1980. He is a Diplomate of the American Board of Toxicology and the American Board of Veterinary Toxicology and a Fellow of the Academy of Toxicological Sciences, Society for Risk Analysis and American Association for Aerosol Research.

He served as Chief Executive Officer and President of the Chemical Industry Institute of Toxicology (CIIT) in Research Triangle Park, NC from September 1988 through July 1999. The CIIT continues today as The Hamner Institute for Health Sciences. During his tenure, the organization achieved international recognition for the development of scientific information under-girding important environmental and occupational health decisions and regulations. Prior to his appointment as President of CIIT, Dr. McClellan was Director of the Inhalation Toxicology Research Institute, and President and Chief Executive Officer of the Lovelace Biomedical and Environmental Research Institute, Albuquerque, New Mexico. The Institute continues operation today as a core element of the Lovelace Respiratory Research Institute. During his 22 years with the Lovelace organization, he provided leadership for development of one of the world's leading research programs concerned with the health effects of airborne radioactive and chemical materials. Prior to joining the Lovelace organization, he was a scientist with the Division of Biology and Medicine, U.S. Atomic Energy Commission, Washington, DC (1965-1966), and Hanford Laboratories, General Electric Company, Richland, WA (1959-1964). In these assignments, he was involved in conducting and managing research directed toward understanding the human health risks of internally deposited radionuclides.

Dr. McClellan is an internationally recognized authority in the fields of inhalation toxicology, aerosol science and human health risk analysis. He has authored or co-authored over 300 scientific papers and reports and edited 10 books. In addition, he frequently speaks on risk assessment and air pollution issues in the United States and abroad. He is active in the affairs of a number of professional organizations, including past service as President of the Society of Toxicology and the American Association for Aerosol Research. He serves in an editorial role for a number of journals, including service since 1987 as Editor of Critical Reviews in Toxicology. He serves or has served on the Adjunct Faculty of 8 universities.

Dr. McClellan has served in an advisory role to numerous public and private organizations. He has served on senior advisory committees for the major federal agencies concerned with human health. This included services as past Chairman of the Clean Air Scientific Advisory Committee, Environmental Health Committee, Research Strategies Advisory Committee, and Member of the Executive Committee, Science Advisory Board, U. S. Environmental Protection Agency; Member, National Council on Radiation Protection and Measurements; Member, Advisory Council for Center for Risk Management, Resources for the Future; a former Member, Health Research Committee, Health Effects Institute; and service on National Academy of Sciences/National Research Council Committees on Toxicology (served as Chairman for 7 years), Risk Assessment for Hazardous Air Pollutants, Health Risks of Exposure to Radon, Research Priorities for Airborne Particulate Matter, as well as the Committee on Environmental Justice of the Institute of Medicine. He has recently completed a term on the Board of Scientific Councilors for the Centers for Disease Control and Prevention for Environmental Health Research and the Agency for Toxic Substances and Disease Registry and on the National Institutes of Health Scientific Advisory Committee on Alternative Toxicological Methods. He currently serves on the National Aeronautics and Space Administration Lunar Airborne Dust Toxicity Advisory Group.

Dr. McClellan's contributions have been recognized by receipt of a number of honors, including election in 1990 to membership in the Institute of Medicine of the National Academy of Sciences. He is a Fellow of the Society for Risk Analysis, the American Association for Aerosol Research, the Health Physics Society, and the American Association for the Advancement of Science. In 1998, he received the International Achievement Award of the International Society of Regulatory Toxicology and Pharmacology for outstanding contributions to improving the science used for decision making and the International Aerosol Fellow Award of the International Aerosol Research Assembly for outstanding contributions to aerosol science and technology. In 2002, he was inducted into the University of New Mexico Anderson School of Management Hall of Fame for contributions to the effective management of multi-disciplinary research organizations. He received the Society of Toxicology Merit Award in 2003 for a distinguished career in toxicology and the Society's Founders Award in 2009 for contributions to science-based safety/risk decision-making. In 2005, The Ohio State University awarded him an Honorary Doctor of Science degree for his contributions to the science under-girding improved air quality. In 2006, he received the New Mexico Distinguished Public Service Award. In 2008, Washington State University presented Dr. McClellan the Regents Distinguished Alumnus Award, the highest recognition the University can bestow on an Alumnus.

Dr. McClellan has a long-standing interest in environmental and occupational health issues, especially those involving risk assessment, air pollution, and safety evaluations for consumer products and pharmaceuticals, and in the management of multidisciplinary research organizations. He is a strong advocate of science-based decision-making and the need to integrate data from epidemiological, controlled clinical, laboratory animal and cell studies to assess human health risks of exposure to airborne materials and establish the safety of consumer products and pharmaceuticals.

COVER SHEET

The proof copy of the manuscript entitled “Critical Considerations in Evaluating Scientific Evidence of Health Effects of Ambient Ozone: A Conference Report” by McClellan, R.O. et al. follows this cover sheet. The manuscript will be published in Inhalation Toxicology 21 (Suppl. 2): 1-36, 2009. This manuscript will be available on-line in September 2009.

Inhalation Toxicology, 2009; 00(00): 000–000

informa
healthcare

REVIEW ARTICLE

Critical considerations in evaluating scientific evidence of health effects of ambient ozone: a conference report

Roger O. McClellan[1], Mark W. Frampton[2], Petros Koutrakis[3], William F. McDonnell[4], Suresh Moolgavkar[5], D. Warner North[6], Anne E. Smith[7], Richard L. Smith[8], and Mark J. Utell[2]

[1]Toxicology and Human Health Risk Analysis, Albuquerque, New Mexico, USA, [2]University of Rochester Medical Center, Rochester, New York, USA, [3]Harvard University School of Public Health, Cambridge, Massachusetts, USA, [4]William F. McDonnell Consulting, Chapel Hill, North Carolina, USA, [5]Exponent, Inc., Bellevue, Washington, USA, [6]NorthWorks, Inc., Belmont, California, USA, [7]CRA International, Washington, DC, USA, and [8]University of North Carolina, Chapel Hill, North Carolina, USA

Abstract

The U.S. Environmental Protection Agency (EPA), under the authority of the Clean Air Act (CAA), is required to promulgate National Ambient Air Quality Standards (NAAQSs) for criteria air pollutants, including ozone. Each NAAQS includes a primary health-based standard and a secondary or welfare-based standard. This paper considers only the science used for revision of the primary standard for ozone in 2008. This paper summarizes deliberations of a small group of scientists who met in June 2007 to review the scientific information informing the EPA Administrator's proposed revision of the 1997 standard. The Panel recognized that there is no scientific methodology that, in the absence of judgment, can define the precise numerical level, related averaging time, and statistical form of the NAAQS. The selection of these elements of the NAAQS involves policy judgments that should be informed by scientific information and analyses. Thus, the Panel members did not feel it appropriate to offer either their individual or collective judgment on the specific numerical level of the NAAQS for ozone. The Panel deliberations focused on the scientific data available on the health effects of exposure to ambient concentrations of ozone, controlled ozone exposure studies with human volunteers, long-term epidemiological studies, time-series epidemiological studies, human panel studies, and toxicological investigations. The deliberations also dealt with the issue of background levels of ozone of nonanthropogenic origin and issues involved with conducting formal risk assessments of the health impacts of current and prospective levels of ambient ozone. The scientific issues that were central to the EPA Administrator's 2008 revision of the NAAQS for ozone will undoubtedly also be critical to the next review of the ozone standard. That review should begin very soon if it is to be completed within the 5-year cycle specified in the CAA. It is hoped that this Report will stimulate discussion of these scientific issues, conduct of additional research, and conduct of new analyses that will provide an improved scientific basis for the policy judgment that will have to be made by a future EPA Administrator in considering potential revision of the ozone standard.

Keywords:

AQ1

Contents

This article is the report of a Working Conference held in Rochester, New York, USA, June 5–6, 2007. The authors are the conference participants.
Address for correspondence: Roger O. McClellan, Advisor, Toxicology and Human Health Risk Analysis, Albuquerque, NM 87111, USA. Fax: 505-296-9573; E-mail: roger.o.mcclellan@att.net

(Received 10 April 2009; revised 09 July 2009; accepted 09 July 2009)

ISSN 0895-8378 print/ISSN 1091-7691 online © 2009 Informa UK Ltd
DOI: 10.1080/08958370903176735

Introduction

This report summarizes the deliberations of the participants in a Working Conference to discuss key considerations in evaluating the scientific evidence on the health effects of ambient ozone that are germane to policy judgments on the setting of the National Ambient Air Quality Standard for ozone. The workshop was organized and participants invited by Mark J. Utell, Professor of Medicine, University of Rochester, with financial support and input from the American Petroleum Institute (API). Participants received copies of the key U.S. Environmental Protection Agency (EPA) documents and reprints of key papers prior to the meeting and identified additional reprints that were distributed to the Panel in advance of the meeting.

The workshop, held June 5–6, 2007 in Rochester, New York, took advantage of the small number of participants to hold discussions in a roundtable format, with individuals leading the conversation in their particular areas of interest. A representative from the API and another from the EPA served as valuable resources and provided clarifying information when requested by the participants of the Working Conference. The API and EPA representatives had no role in preparation or review of the report. Roger McClellan agreed to coordinate the preparation of a summary report based on written material submitted by the participants. All participants reviewed and commented on the entire report to help ensure its accuracy and clarity for a broad audience. However, readers of the report should recognize that although the individual participants are experts in one or more of the topics covered, no participant had an in-depth knowledge of all the areas covered. In areas as complex as those covered at the Conference and in this report, individual scientists may have differing views on interpretation of specific scientific issues. Moreover, individual scientists reviewing the same science may reach different judgments in applying the science in the standard-setting process. Thus, the participants did not attempt to forge a consensus on all scientific issues. Most importantly, the participants did not offer either individual judgments or a consensus judgment as to the numerical level of the National Ambient Air Quality Standard (NAAQS) for ozone. The participants agreed that although the science reviewed and discussed should inform the setting of the NAAQS for ozone, the selection of a specific numerical level, averaging time, and associated statistical form for the NAAQS was not a specific focus of the workshop and, most importantly, such policy judgments are ultimately the responsibility of the EPA Administrator as described in the next section.

Clean Air Act authority for setting NAAQS

Section 108 (42 U.S.C. § 7408) directs the Administrator to identify and list "air pollutants" that "in his judgment, may reasonably be anticipated to endanger public health and welfare" and whose "presence…in the ambient air results from numerous or diverse mobile or stationary sources" and to issue air quality criteria for those that are listed. Air quality criteria are intended to "accurately reflect the latest scientific knowledge useful in indicating the kind and extent of identifiable effects on public health or welfare

which may be expected from the presence of [a] pollutant in ambient air....".

Section 109 (42 U.S.C. § 7409) directs the Administrator to propose and promulgate "primary" and "secondary" NAAQS for pollutants listed under section 108. Section 109(b)(1) defines a primary standard as one "the attainment and maintenance of which in the judgment of the Administrator, based on such criteria and allowing an adequate margin of safety, are requisite to protect the public health." The legislative history of § 109 indicates that a primary standard is to be set at "the maximum permissible ambient air level...which will protect the health of any [sensitive] group of the population," and that for this purpose "reference should be made to a representative sample of persons comprising the sensitive group rather than to a single person in such a group" (S. Rep. No. 91-1196, 91st Cong., 2d Sess. 10 (1970).

A secondary standard, as defined in § 109(b)(2), must "specify a level of air quality the attainment and maintenance of which, in the judgment of the Administrator, based on such criteria, is requisite to protect the public welfare from any known or anticipated adverse effects associated with the presence of [the] pollutant in the ambient air." Welfare effects as defined in § 302(h) (42 U.S.C. § 7602(h)) include, but are not limited to, "effects on soils, water, crops, vegetation, man-made materials, animals, wildlife, weather, visibility and climate, damage to and deterioration of property, and hazards to transportation, as well as effects on economic values and on personal comfort and well-being." This report deals exclusively with the scientific issues involved in setting the primary (public health) standard and does not consider issues related to the setting of the secondary (welfare) standard.

The requirement that primary standards include an adequate margin of safety was intended to address uncertainties associated with inconclusive scientific and technical information available at the time of standard setting. It was also intended to provide a reasonable degree of protection against hazards that research has not yet identified. [*Lead Industries Association v. EPA*, 647 F.2d 1130, 1154 (D.C. Cir 1980), *cert. denied*, 449 U.S. 1042 (1980); *American Petroleum Institute v. Costle*, 665 F.2d 1176, 1186 (D.C. Cir. 1981), *cert. denied*, 455 U.S. 1034 (1982).] Both kinds of uncertainties are components of the risk associated with pollution at levels below those at which human health effects can be said to occur with reasonable scientific certainty. Thus, in selecting primary standards that include an adequate margin of safety, the Administrator is seeking not only to prevent pollution levels that have been demonstrated to be harmful but also to prevent lower pollutant levels that may pose an unacceptable risk of harm, even if the risk is not precisely identified as to nature or degree. The Clean Air Act does not require the Administrator to establish a primary NAAQS at a zero-risk level or at background concentration levels (*see Lead Industries Association v. EPA*, 647 F.2d at 1156 n.51), but rather at a level that reduces risk sufficiently so as to protect public health with an adequate margin of safety.

In addressing the requirement for an adequate margin of safety, EPA considers such factors as the nature and severity of the health effects, the size of the population(s) at risk, and the nature and degree of uncertainties that must be addressed. The selection of any particular approach to providing an adequate margin of safety is a policy choice left specifically to the Administrator's judgment. *Lead Industries Association v. EPA*, 647 F.2d at 1161-62; Whitman v. American Trucking Associations, 531 U.S. 457, 495 (2001) (Breyer, J., concurring in part and concurring in judgment).

In setting standards that are "requisite" to protect public health and welfare, as provided in section 109(b), EPA's task is to establish standards that are neither more or less stringent than necessary for these purposes (*Whitman v. American Trucking Associations*, 531 U.S. 457, 473). In establishing "requisite" primary and secondary standards, EPA may not consider the costs of implementing the standards (*Id.* at 471). As discussed by Justice Breyer in *Whitman v. American Trucking Associations*, however, "this interpretation of § 109 does not require the EPA to eliminate every health risk, however slight, at any economic cost, however great, to the point of "hurtling" industry over "the brink of ruin," or even forcing "deindustrialization." (*Id.* at 494) (Breyer, J., concurring in part and concurring in judgment) (*citations omitted).* Rather, as Justice Breyer explained:

> *"The statute, by its express terms, does not compel the elimination of all risk; and it grants the Administrator sufficient flexibility to avoid setting ambient air quality standards ruinous to industry.*
> *Section 109(b)(1) directs the Administrator to set standards that are "requisite to protect the public health" with "an adequate margin of safety." But these words do not describe a world that is free of all risk—an impossible and undesirable objective. (citation omitted). Nor are the words "requisite" and "public health" to be understood independent of context. We consider football equipment "safe" even if its use entails a level of risk that would make drinking water "unsafe" for consumption. And what counts as "requisite" to protecting the public health will similarly vary with background circumstances, such as the public's ordinary tolerance of the particular health risk in the particular context at issue. The Administrator can consider such background circumstances when "deciding what risks are acceptable in the world in which we live." (citation omitted).*
>
> *The statute also permits the Administrator to take account of comparative health risks. That is to say, she may consider whether a proposed rule promotes safety overall. A rule likely to cause more harm to health than it prevents is not a rule that is "requisite to protect the public health." For example, as the Court of Appeals held and the parties do not contest, the Administrator has the authority to determine to what extent possible health risks stemming from reductions in tropospheric ozone (which, it is claimed, helps prevent cataracts and skin cancer) should be taken*

into account in setting the ambient air quality standard for ozone. (Citation omitted)/

The statute ultimately specifies that the standard set must be "requisite to protect the public health" "in the judgment of the Administrator," § 109(b)(1), 84 Stat. 1680 (emphasis added), a phrase that grants the Administrator considerable discretionary standard-setting authority.

The statute's words, then, authorize the Administrator to consider the severity of a pollutant's potential adverse health effects, the number of those likely to be affected, the distribution of the adverse effects, and the uncertainties surrounding each estimate. (citation omitted). They permit the Administrator to take account of comparative health consequences. They allow her to take account of context when determining the acceptability of small risks to health. And they give her considerable discretion when she does so.

This discretion would seem sufficient to avoid the extreme results that some of the industry parties fear. After all, the EPA, in setting standards that "protect the public health" with "an adequate margin of safety," retains discretionary authority to avoid regulating risks that it reasonably concludes are trivial in context. Nor need regulation lead to deindustrialization. Pre-industrial society was not a very healthy society; hence a standard demanding the return of the Stone Age would not prove "requisite to protect the public health."

Although I rely more heavily than does the Court upon legislative history and alternative sources of statutory flexibility, I reach the same ultimate conclusion. Section 109 does not delegate to the EPA authority to base the national ambient air quality standards, in whole or in part, upon the economic costs of compliance."

Section 109(d)(1) of the Clean Air Act requires that "not later than December 31, 1980, and at 5-year intervals thereafter, the Administrator shall complete a thorough review of the criteria published under section 108 and the national ambient air quality standards… and shall make such revisions in such criteria and standards and promulgate such new standards as may be appropriate…." Section 109(d)(2) requires that an independent scientific review committee "shall complete a review of the criteria…and the national primary and secondary ambient air quality standards…and shall recommend to the Administrator any new…standards and revisions of existing criteria and standards as may be appropriate…" This independent review function is performed by the Clean Air Scientific Advisory Committee (CASAC) of EPA's Science Advisory Board.

Historical review of the NAAQS for ozone

Each NAAQS consists of four elements: an indicator, an averaging time, a numerical level, and a statistical form. The initial NAAQS for photochemical oxidants was promulgated on April 30, 1971, with both the primary and secondary standard set at 0.08 ppm, total photochemical oxidants, not to be exceeded more than 1 h per year. On February 8, 1977, the NAAQS for photochemical oxidants was revised. The indicator was changed to ozone, and the associated analytical method changed, with the level of both the primary and secondary standard set at 0.12 ppm ozone for a 1-h averaging time. The form of the standard was also changed to one based on the expected number of days per calendar year with a maximum hourly average concentration above 0.12 ppm (i.e., attainment of the standard occurs when that number is equal to or less than 1). On July 18, 1997, the ozone NAAQS was revised with the averaging time changed from 1 to 8 h and the numerical level set at 0.08 ppm, which may be viewed as equivalent to 0.084 ppm using the standard rounding convention. The form of the standard was changed to the annual fourth-highest daily maximum 8-h average concentration, averaged over 3 years. Without going into the details, provision was made for an orderly transition from the 1-h averaging time to an 8-h averaging time. It is obvious that 1-h averaging time values exceed 8-h averaging time ozone values, which in turn exceed 24-h averaging time ozone values. The relationship between the three averaging time values varies day to day, throughout the year, and among different communities. In 1996–1997, it was viewed on average that a 0.12-ppm, 1-h averaging time value, was approximately equivalent to 0.09 ppm ozone averaged over 8 h. Thus, the 1997 NAAQS of 0.08 ppm averaged over 8 h was expected to be somewhat more restrictive than the previous 1-h averaging time NAAQS set at 0.12 ppm ozone. The 1997 revision of the NAAQS for ozone was based on the EPA Criteria Document (1996a) and Staff Paper (1996b).

The EPA initiated, in September 2000, the review completed in March 2008 with a call for information for the development of a revised Air Quality Criteria Document for ozone and other photochemical oxidants. The EPA, under a court decree, after several extensions, was required to complete its review of the scientific criteria for ambient ozone limits and sign the publication notices of proposal and final rule making of the ozone NAAQS by June 20, 2007, and March 12, 2008, respectively. To meet this schedule, the EPA prepared a new Criteria Document (2006) and Staff Paper (2007a). The Staff Paper was supported by an ozone population exposure analysis for selected urban areas (EPA, 2007b), an analysis of uncertainties in the exposure analysis (EPA, 2007c) and a detailed ozone health risk assessment for selected urban areas (EPA, 2007d).

The Administrator met the June 20, 2007, deadline for the proposed rule making with the Proposed Rule published in the July 11, 2007 Federal Register (EPA, 2007e). In summary, "The Administrator's proposed decision is to revise the existing 8-h ozone primary standard by lowering the level to within a range from 0.070 to 0.075 ppm, and to specify the standard to the nearest thousandths ppm (i.e., to the nearest parts per billion)." The Proposed Rule, "the EPA solicits comments on alternative levels down to 0.060 ppm and up to and including retaining the current 8-h standard of 0.08 ppm (effectively 0.084 ppm using current

AQ2 data rounding conventions)." The EPA's proposed rule was based on consideration of evidence of ozone health effects from controlled human exposure, epidemiological and toxicological studies. In this report, each of these kinds of evidence is discussed and critiqued. By way of background, the EPA also prepares an additional document, a Regulatory Impact Analysis, to support the development of air pollution regulations. That report (EPA, 2007f) was released on August 2, 2007. The report is not part of the standard selection process. It is intended to inform the public and States about the potential costs and benefits of implementing the proposed air quality standards. The Panel did not review that report. However, it is obvious that many of the issues discussed in this report, especially those addressing risk assessment issues, are also relevant to the preparation and interpretation of the Regulatory Impact Analysis.

EPA Administrator Stephen Johnson announced the Final Rule for the National Ambient Air Quality Standard for ozone on March 12, 2008, and the Final Rule was published in the March 27, 2008, Federal Register (EPA, 2008). The NAAQS for ozone was revised with the primary standard set at 0.075 ppm for an 8-h averaging time. The statistical form of the standard was not changed, namely, the revised 8-h primary standard of 0.075 ppm will be met at an ambient monitoring site when the 3-year average of the annual four highest daily maximum 8-h average O_3 concentration is less than or equal to 0.075 ppm.

Framework for evaluating health effects

The general framework used for integrating scientific information on ambient air pollutants and their health effects is shown in Figure 1. This framework has been widely used in evaluating the health risks of airborne materials and provided the basis for the deliberations and recommendations in the four reports of the National Research Council (NRC) Committee on "Research Priorities for Airborne Particulate Matter" (NRC, 1998, 1999, 2001, 2003). We refer to it here to make several overarching points.

Public health goal

It is important to emphasize that explicit in the Clean Air Act is the intent to regulate air pollutants, such as the criteria air pollutants including ozone, to manage health impacts of air pollution to acceptable risk levels. This is accomplished in a multistep process. The first is to establish NAAQS for specific pollutants such as ozone. In a second step, measures are taken to limit the man-made emissions of the pollutant or its precursors to achieve ambient air concentrations compliant with the appropriate NAAQS.

Ozone is formed in the troposphere from precursors that are of both biogenic and anthropogenic origins. Thus, even in the absence of man-made (anthropogenic) emissions of ozone precursors, there are ambient concentrations of ozone from biogenic precursors, lightning, and periodic intrusions of stratospheric ozone into the troposphere. To assist in considering the portion of the ambient ozone of natural origin, and, thus, not readily controllable, the EPA has introduced the concept of "Policy Relevant Background" ozone. It is defined "as the ozone concentrations that would be observed in the U.S. in the absence of anthropogenic emissions of precursors (e.g., VOC, NO_2 and CO) in the U.S., Canada and Mexico." The issue of Policy Relevant Background ozone is of sufficient importance that it will be discussed in a separate section of this paper.

Monitoring data

Substantial monitoring data on regulated pollutants are routinely collected in locales across the United States, primarily to demonstrate compliance with the NAAQS for the individual air pollutants. Other pollutants are monitored much less frequently and typically on a "campaign" basis rather than the year-after-year, decade-after-decade approach taken with the Criteria Pollutants. Ozone levels have generally been decreasing in most communities across the United States as a result of air quality control measures, with an estimated 21% reduction (8-h maximum average) and a 29%

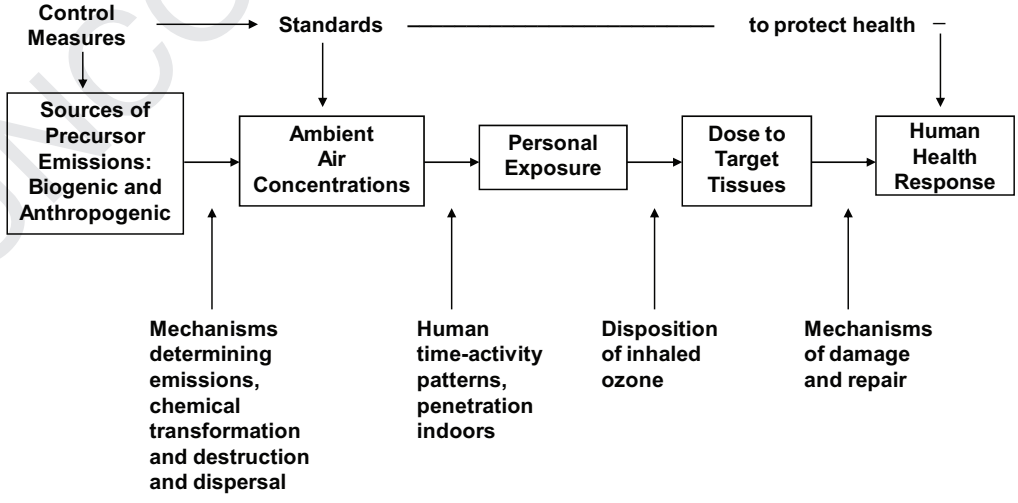

Figure 1. Framework for evaluating health risks of ozone.

reduction (1-h maximum average) in ambient ozone nationwide since 1980 (EPA, 2007g).

Significant progress has also been made in reducing ambient concentrations of other criteria pollutants. In addition to evaluating compliance with the NAAQS, the routine monitoring data have also been used extensively as indices of exposure in epidemiological studies in which the air quality data are most frequently used in combination with data on morbidity and mortality of populations obtained from administrative databases.

Personal exposure

It is obvious that individuals breathe whatever is in the air in their personal breathing zone. The concentration of ozone measured at monitoring stations is an imperfect surrogate for the ozone concentration at other locations. The correlation between ambient monitor concentrations and that found in the interiors of homes, offices, other workplaces, and other buildings is poor because of the high reactivity of ozone, deposition on surfaces, and the influence of building ventilation. As discussed in a later section, some studies have attempted to characterize the actual exposure of individuals. Those kinds of exposure data are valuable for studying small groups of individuals. However, it is not feasible to acquire personal exposure data on large populations of individuals over long periods of time.

Dosimetry of inhaled ozone

Substantial knowledge has been acquired of the relationship between quantities of ozone inhaled and the deposition of ozone in the respiratory tract of humans and some laboratory animal species. This information is of special value in understanding how inhaled ozone affects the body. Moreover, it is useful in understanding how toxicological data acquired in laboratory animal species may be extrapolated to humans, recognizing critical species differences in ozone dosimetry.

Conduct of epidemiological investigations

Epidemiological investigations provide highly relevant evidence as to whether ambient environmental factors, such as air pollutants, including ozone, can adversely affect the health of the general public. These studies are especially valuable because they consider the real-world experiences of people as they are exposed to air pollution in the course of their daily lives. The majority of the studies of ozone exposure have been observational semiecological studies that analyze aggregate ambient concentration data as a surrogate for exposure for large groups of people. The studies seek to evaluate whether a statistical association exists between ambient ozone concentrations and a health outcome. By assuming that ambient concentrations are correlated with exposures, and exposures are correlated with personal doses, epidemiological studies attempt to link ambient ozone and health outcomes shown in Figure 1. Ozone concentrations measured at one or a few central monitors are used as an index of exposure. As noted earlier, the monitoring data are typically required for regulatory compliance purpose; their use in epidemiological studies is a fringe benefit. The monitors are rarely located with special attention given to the location as being representative of a specific population. The ozone data may be available from daily measurements made year round or only in the summer when ozone concentrations are generally higher. They may be aggregated as a daily 24-h average, the highest 8-h average or the maximum 1-h concentration. The latter two indices reflect past NAAQSs for ozone that used 8-h or 1-h averaging times. At many monitoring stations, data are acquired on other criteria pollutants, temperature, and relative humidity. Various health indices, such as measures of morbidity or mortality, are evaluated for their association with ambient ozone concentration. The result may be expressed as a proportional increase or decrease in the health index per increment of ozone concentration. None of the health effects evaluated are unique to ozone. Exposure to other pollutants, temperature, relative humidity, other indicators of weather, as well as population characteristics such as age, health status, socioeconomic status, housing, and exercise also influence the various health effects. The role of ozone must be statistically separated from these other health determinants, which presents the most difficult challenge in epidemiological studies of this sort.

Overarching issues

Policy Relevant Background

The issue of defining and then characterizing the background concentrations of ozone is of critical importance for several reasons to any potential revision of the NAAQS for ozone. The definition used for background ozone will determine the relevant background ambient ozone concentrations. In this NAAQS ozone review, the EPA has coined the term, Policy Relevant Background (PRB) of ozone. This is defined as the ozone concentrations that would be observed in the United States in the absence of anthropogenic emissions of precursors (e.g., VOC, NO_2, and CO) in the United States, Canada, and Mexico. Indeed, the Policy Relevant Background values used in the Staff Paper and Risk Assessment were set by "turning off the inputs for man-made precursors of ozone" in a low-resolution chemical transport model (GEOS-CHEM) (Fiore et al., 2002, 2003). The GEOS-CHEM model has very low spatial resolution (2 degrees by 2.5 degrees). This corresponds to 138 miles by 173 miles with an area of 24,000 square miles, an area about half that of the Commonwealth of Pennsylvania.

It is especially significant that the manner in which the model was used zeros out all Canadian and Mexican man-made emissions of ozone precursors. As a result, the Policy Relevant Background estimates the EPA uses overstate the levels of ozone that U.S. regulations can potentially control. An alternative definition of Policy Relevant Background for ozone would be the background ambient concentrations of ozone projected for the United States in the absence of emissions of man-made precursors of ozone in the United States. Using this definition would allow characterization of the concentrations of ozone attributable to man-made precursors

by the difference between measured or modeled ambient ozone concentrations and the estimated background concentrations of ozone. This would be the portion of ambient ozone that could be influenced by U.S. regulations and the actions of the regulated sources of ozone precursors.

Getting the background ambient concentrations of ozone correct across the United States is vitally important to understanding how the background concentration of ozone will influence the likelihood that communities will attain any particular NAAQS for ozone. Since 1997, the NAAQS for ozone has had an averaging time of 8 h and an associated statistical form. Thus, it is necessary to characterize the distribution of each hour's ozone PRB over multiple days to capture the variability in PRB and not just a typical or "average" level in order to determine how the background levels of ozone will influence the attainment of the 8-h averaging time. With a robust statistical form used to define attainment of the 8-h averaging time standard, infrequent high levels of background ozone may keep many communities from attaining the standard, despite stringent limits on the emissions of man-made precursors of ozone. The lower the numerical level of the 8-h averaging time standard, the higher the likelihood that background levels of ozone will influence whether the community is in or out of attainment. This likelihood cannot be captured in modeled estimates of PRB that do not adequately reflect its day-to-day variability and spatial variability.

How background concentrations of ozone are defined and characterized has a major influence on the results of the risk assessment for current levels of ozone and various potential levels of NAAQS for ozone and the interpretation of these results. By assuming what are likely to be unrealistically low estimates of background ozone, the EPA has calculated risks at low ambient ozone concentrations that are not controllable with U.S. regulations and, further, treated these calculated risks as though they were attributable to ozone arising from man-made sources in the United States.

The EPA, in its proposed rule for ozone, acknowledged the shortcoming of the treatment of "Policy Relevant Background" in the Staff Paper, the Risk Assessment, and the Proposed Rule by noting in footnote 40 to the Proposed Rule that it will perform further analyses on this issue (EPA, 2007f). A good starting point for such analyses would be to include in a future Criteria Document and Staff Paper alternative approaches to considering background concentrations of ozone. Papers on ozone background levels that deserve careful consideration include those of Vingarzan (2004) and Oltmans et al. (2006). The importance of the issue of background levels of ozone is such that any revised information developed on this issue should be open for public review and comment. It is vitally important to recognize that choices made in estimating background ozone levels will necessitate changes in the risk assessment for ozone as regards the portion of estimated risk that may be attributed to background levels of ozone versus that attributed to man-made sources.

Variability and uncertainty

The topics of uncertainty and variability in the context of environmental risk assessment and regulatory decision making have been discussed at length in many reports of the National Research Council (NRC, 1983, 1993, 1994, 2002, 2007a, 2007b) prepared over the past 25 years. This section briefly summarizes definitions and main themes. For more extensive discussion the reader is referred to these National Academies reports and the multitude of references they contain.

The basic problem is a lack of data and a lack of scientific understanding and predictability of the impact of regulation on improving human health and the environment. In the transmittal letter for the seminal 1983 NRC report (NRC, 1983, page iii), National Academy of Science President Frank Press stated, " … the committee finds that the basic problem in risk assessment is the incompleteness of data … ." The Summary of that report provides an expanded version of this statement by the authoring Committee: "The Committee believes that the basic problem in risk assessment is the sparseness and uncertainty of the scientific knowledge of the health hazards addressed, and this problem has no ready solution" (NRC, 1983, page 6). This theme is reiterated in the 1994 report, citing statements from the 1983 report similar to the two quotations above (NRC, 1994, page 160).

Uncertainty

"Uncertainty can be defined as a lack of precise knowledge as to what the truth is, whether qualitative or quantitative. That lack of knowledge creates an intellectual problem—that we do not know what 'scientific truth' is; and a practical problem—we need to determine how to assess and deal with risk in the light of that uncertainty" (NRC, 1994, page 161). "Scientific truth is always somewhat uncertain and is subject to revision as new understanding develops, but uncertainty in quantitative health risk assessment might be uniquely large, relative to other science policy areas, and it requires special attention by risk analysts." (Ibid. See also the Summary [NRC, 1994, pages 11–12]). In the context of this document, uncertainty refers to lack of knowledge on the human health impacts of changes in exposure to ozone and associated air pollutants for which monitored ozone concentrations may be an indicator. Uncertainty may also refer to lack of knowledge in the relation of ambient ozone levels to precursor emissions.

Variability

Variability may be defined as a description of differences among a population or a set of situations that one can describe in applying classical statistical methods ~~apply~~ to describing these differences. Variation in people's weight provides a simple example. A distribution of weights of adult males in the United States population describes the probability that the weight of an adult male chosen at random, among all those in the United States, will fall within given intervals from the lightest to the heaviest. For example, the probability of a weight of 150 to 155 pounds might be given

as 4.1%, signifying that 4.1% of American adult males fall within this interval and that 95.9% weigh less than 150 or more than 155 pounds. If a subpopulation is selected, such as adult males from 60 up to 65 years of age, then this probability may change. Now only a subset of the data on weight of adult males, those aged 60–64, rather than all males at least 21 years of age, are used for the distribution.

A more complex example is susceptibility to health impact, such as a change in lung function measurement (such as forced expiratory volume at 1 second [FEV_1]) or in recorded symptoms, such as pain on deep inspiration resulting from exposure while exercising to ozone for a specific length of time and at a specific concentration. This is not a situation brought about by lack of data, but rather that the available data indicate that individuals differ in their response to an air pollutant such as ozone. For example, the subset of people who have been diagnosed with asthma may differ from those who have not been so diagnosed in their response to ozone. Similarly, because of differences in meteorological conditions and other factors, there is a great deal of variability how reductions in emissions of ozone precursors affect the ozone concentration levels that different individuals with differing exercise patterns in different locations experience. A detailed statistical description of ozone exposure over time to a population of individuals in different locations in a major city can become quite complex, even if all the relationships involved were known.

Both NRC Reports (NRC, 1983, 1994) urge EPA to improve its characterization of uncertainty in risk assessments, and to disaggregate and justify its assumptions in dealing with uncertainty (limitations in scientific knowledge and lack of data) and in variability among individual humans and exposure situations. Both uncertainty and variability can be treated using modern statistical methods, which include the use of expert judgment and inference methods to develop probability distributions in the absence of data. (Good recent reviews on uncertainty in risk assessment are found in NRC [2007a], pages 79–88, and NRC [2007b], pages 43–52.)

Effective communication about uncertainty and variability is needed to inform regulatory decision makers and the public. All of the National Research Council Reports (NRC, 1983, 1993, 1994, 2002, 2007a, 2007b) suggest that quantitative risk assessments can be helpful in achieving such effective communication. EPA needs to improve its practices in carrying out and communicating the results from risk assessments. "To the extent that both uncertainty and inter-individual variability (that is, heterogeneity or differences among people at risk) are addressed quantitatively with separate input components (e.g., ambient concentrations, uptake, and potency) for aggregation into an assessment of risk, the distinction between uncertainty and variability ought to be maintained rigorously throughout the analytic process, so that uncertainty and variability can be distinctly reflected in calculated risk" (NRC, 1994, page 239; see also the two summary points on page 242).

Because of the multitude of factors involving uncertainty, variability, and often both, simplifying assumptions are made to carry out risk assessments. Such assumptions typically involve value judgments, and such judgments are often made so as to be conservative in the protection of human health. For example, health effects might be computed assuming exposure at a location estimated to have the largest cumulative exposure over time to a toxic air contaminant. Health effects for such a hypothetical "maximally exposed individual" (MEI) should be greater than for people who are at other locations, and even for people who may spend some time at the point of maximum exposure but not all of their time. The analysis using an MEI is greatly simplified from that of calculating a distribution of time-varying exposures and consequent estimated health impacts. As described in NRC (NRC, 1994), especially Chapter 10, simplified analysis may be appropriate for screening calculations that investigate whether potential health impacts are large enough to warrant further investigation and regulatory decision making. For important assessments of health impacts, such as those for ozone, EPA appropriately uses much more complex analytical methods for ~~describe~~ exposures over time and by location, but these still require significant amounts of simplifying policy judgments.

Simplifying policy judgments that deal with uncertainty and variability are referred to in NRC Report (1994) as "inference guidelines" and in NRC Report (1994) as "defaults." Such judgments are needed for risk assessments, because without such simplification, analysis of most situations involving exposure to pollutants becomes impractically complicated. Therefore, sensitivity analysis and judgment are used to determine where uncertainty and variability should be explicitly addressed. Both NRC Reports (1983, 1994) urge a clarification in EPA practices through the use of guidelines and specific criteria for the use of defaults, or for departure from a default when justified by newly available scientific information. Appendices N-1 and N-2 of NRC (NRC, 1994) present contrasting viewpoints on appropriate criteria for selection of defaults and departure from defaults. Although the details are complex, both of these appendices and the main report of NRC (NRC, 1994) urge EPA to improve its use of quantitative analysis and resulting characterization of uncertainty and variability, for regulatory decision makers and for the public.

The 2002 NRC report (2002) addressed the specific problem of estimating public health benefits from regulation of emission precursors to pollutants such as ozone and fine particulate matter through NAAQS. The report describes concerns that estimates of morbidity and mortality—"body counts"—are being used in regulatory impact analysis without sufficient use of probabilistic methods to describe uncertainty and variability. "EPA should begin to move the assessment of uncertainties from its ancillary analyses into its primary analyses by conducting probabilistic, multiple-source uncertainty analysis. This shift will require specification of probability distributions for major sources of uncertainty. These distributions should be based on available data and expert judgment" (NRC, 2002, page 14). (Further discussion is found on pages 125–152.). "Although the results of benefit analysis may appear to be less certain, EPA

should describe the uncertainty as completely and realistically as possible, recognizing that regulatory action might be necessary in the presence of substantial uncertainty" (NRC, 2002, page 15). "… uncertainties in each stage of the analysis should be quantified and carried through the entire [health benefits analysis] process" (NRC, 2002, page 153).

EPA has not, to date, incorporated uncertainty analysis into its primary risk analyses and we find that the overall interpretation of the risk analysis for ozone is potentially misleading as a result. Considerable uncertainty remains on the impact of reducing the 8-h ozone standard from the current level of 0.08 ppm to a lower level. EPA has assumed for its analysis that shifts in the probability distribution of exposure to ozone will occur not just in the upper tail of the distribution (the days on which ozone levels will exceed the existing 8-h standard of 0.08 ppm or proposed alternative standards of 0.060–0.074 ppm), but also the lower and middle portions of this distribution, levels below 0.060 ppm. It is these changes in the lower and middle portions of the distribution that provide most of the calculated benefits in terms of reduced short term-mortality and reduced hospital admissions associated with the more stringent alternative proposed standard. These matters are discussed in more detail later when the risk assessment process is discussed.

Exposure assessment

Personal exposures and indoor concentrations

Quite a few panel studies found that ozone personal exposures were substantially lower than ozone outdoor concentrations. These exposure assessment studies were conducted in a variety of environments, including Toronto (Liu et al., 1995); Vancouver (Brauer et al., 1995); Baltimore (Sarnat et al., 2000); Southern California (Geyh et al., 2000); Nashville (Lee et al., 2004); Mexico City (O'Neill et al., 2003); Boston (Koutrakis et al., 2005); and Steubenville (Sarnat et al., 2006). The low ozone personal exposures found in these panel studies are due to the high proportion of time spent indoors by study participants and the high deposition rate of ozone onto the microenvironmental surfaces. For example, Liu et al. (1995) and Brauer and Brook (1995) showed that subjects who spend more time outdoors were exposed to higher ozone levels.

Indoor concentrations of ozone are typically considerably lower than those measured outdoors, with indoor/outdoor ratios depending to a great extent on home ventilation. For homes with low air exchange rates, for example homes using air conditioning, indoor/outdoor ratios are very low. Consequently, in cities with hot summers, only a small fraction of ambient ozone can be found indoors.

In Nashville, Tennessee, summer indoor ozone concentrations ranged from 3% to 15% of outdoor concentrations with an average indoor/outdoor ratio of 0.10.[30] This ratio was lower than that (0.30) measured for two communities during the summer in Southern California (Geyh et al,. 2000). Indoor ozone concentrations are very low during the winter season because homes are tight, especially in cities with harsh winters where home air ventilation rates are lower than one air exchange per hour. Liu et al. (1995) measured indoor ozone concentrations in 50 homes in Toronto, Canada, and found that the average indoor/outdoor ratio was 0.11. A similar ratio was reported by Geyh et al. (2000) in the winter Southern California study. So even in locations without harsh winters, tight homes during this season can result in quite low indoor/outdoor ratios. Romieu et al. (1998) reported home indoor ozone concentrations in Mexico City that were 10%–30% of ambient concentrations with higher indoor concentrations in homes with windows open during the day. Indoor/outdoor ratios in schools, where windows and doors were frequently open, ranged between 0.3 and 0.4 and were higher than those observed homes, which were about 0.2 (1998).

Relationships between ozone personal exposures and outdoor concentrations

In the Baltimore (Sarnat et al., 2000) and Boston (2005) studies, ozone personal exposures were weakly correlated with ozone ambient concentrations. When personal exposures were regressed on ambient concentrations using mixed models, the estimated slopes were substantially lower than unity. In Baltimore, the estimated slopes for the winter and summer seasons were 0.00 (CI: –0.02, 0.02) and 0.04 (CI: –0.02, 0.10), respectively, which were not statistically significant. In Boston, the slopes were 0.05 (CI: 0.02, 0.08) for winter (nonsignificant) and 0.27 (CI: 0.13, 0.39) for summer (significant). Sarnat et al. (2006), for a panel study of elderly individuals in Steubenville, Ohio, found personal exposures were correlated with outdoor concentrations, but indoor ozone concentrations were substantially lower than ozone outdoor concentrations.

Somewhat different findings were reported by O'Neill et al. (2006) who measured exposures of outdoor workers (shoe cleaners) in Mexico City. In this study, strong within-worker longitudinal associations between ambient ozone concentrations and personal exposures were observed. The difference in ambient-personal relationships in this study compared to those presented above, likely relates to the substantial differences in time spent outdoors. This finding suggests that fixed-site ozone monitors may adequately estimate exposures in repeated-measure health studies of outdoor workers.

Relationships between personal $PM_{2.5}$ exposures and outdoor ozone concentration

During the summer season, when outdoor ozone and $PM_{2.5}$ concentrations are correlated, outdoor ozone concentrations may be surrogates for personal $PM_{2.5}$ exposures. This is supported by the Boston, Baltimore, and Steubenville summer panel studies of Koutrakis et al. (2005) and Sarnat et al. (2006). In contrast, the Boston and Baltimore winter panel studies and the Steubenville fall panel study did not show associations between personal $PM_{2.5}$ exposures and outdoor ozone levels. This is because outdoor ozone and $PM_{2.5}$ concentrations were inversely correlated, which is typical for northeastern urban environments during these seasons.

Exposure assessment implications for health effects studies

Panel studies. A number of panel health effects studies have examined associations between respiratory morbidity and ozone exposures. Except for a few studies, including Delfino et al. (1997) and Brauer et al. (1996), the majority of these panel investigations have relied on ambient monitors to estimate human exposures. Considering that several panel exposure studies have shown substantial inter- and intrapersonal variability in personal exposure/ambient concentration ratios, reliance on ambient monitors, especially for large geographic areas and for populations that do not spend a high proportion of the study period outdoors, may be inadequate and can introduce considerable bias. The impact of this exposure error is more severe on panel studies that examine a relatively small number of subjects as compared to that for population studies.

Time-series analyses. In cities with mild summers and/or winters, positive associations were found between ozone exposures and outdoor concentrations. In contrast, for cities with very hot summers and/or harsh winters, population ozone exposures may not be associated with the corresponding ambient concentrations, thus it is difficult to study the effects of ambient ozone. Furthermore, when interpreting the results of the ozone time-series health studies, it is important to keep in mind that during the summer outdoor ozone can be a surrogate of personal $PM_{2.5}$ or other pollutant exposures. Consequently, it may not be possible to distinguish between the health effects of ozone and those of fine particles or other pollutants when using ambient measurements of ozone during the summer season.

Multi-city studies. The extent of exposure error depends on the climatic conditions, and is more pronounced for populations living in cities with very hot summers (e.g., Baltimore and Nashville) and harsh winters (e.g., Toronto and Boston). Therefore, future exposure assessment studies should determine city specific population exposure/ambient concentrations ratios.

CASAC raised issue of exposure error. The importance of exposure error was presented in a letter of the CASAC ozone panel to the EPA Administrator, dated June 05, 2006. Below we present some excerpts regarding ozone exposure assessment issues:

> "The Ozone Staff Paper should consider the problem of exposure measurement error in ozone mortality time-series studies. It is known that personal exposure to ozone is not reflected adequately, and sometimes not at all, by ozone concentrations measured at central monitoring sites. Typically, personal exposures are much lower than the ambient concentrations, and can be dramatically lower depending on the time-activity patterns, housing characteristics and season. In addition, and of particular importance for the ozone time-series studies, there can be no correlation between personal concentrations of ozone measured over time and concentrations measured at central outdoor sites. The population that would be expected to be potentially susceptible to dying from exposure to ozone is likely to have ozone exposures that are at the lower end of the ozone population exposure distribution, in which case this population would be exposed to be very low concentration of ozone indeed, and especially so in winter. Therefore, it seems unlikely that the observed associations between short-term ozone concentrations and daily mortality are due solely to ozone itself."
>
> "Another implication of the ozone measurement error that is relevant in the ozone NAAQS-setting process is that the degree of measurement error would be expected to have a substantial impact on the ability to detect a threshold of the concentration-response relationship below which no ozone effects are discernable. Pollutant exposure measurement error obscures true thresholds in the concentration-response relationship, and this effect worsens with increasing degrees of measurement error. Since threshold assumptions are incorporated in the Agency's risk assessment and risk analyses, this issue will need to be addressed."
>
> "At least two questions arise from these observations that are relevant to the ozone NAAQS-setting process: (1) What chemical agent or agents are at least partly responsible for the observed associations between ozone and mortality in the time-series studies? and (2) Do we require an immediate answer to the question of whether ambient ozone adequately serves as a surrogate marker, that, when controlled, effectively mitigates health impacts of this entire mix of pollutants? One possible explanation for the observed associations of ozone is that ozone itself serves as a marker for other agents that are contributing to the short-term exposure effects on mortality. This would require that outdoor concentrations of these agents are correlated over time with outdoor ozone concentrations, which is to be expected if they are products of the same process that lead to ozone formation, and that these outdoor pollutant concentrations are better correlated with personal exposures than is the case for ozone itself."

We strongly agree with the exposure error issues raised by the CASAC panel, which are based on a thorough review of the ozone exposure assessment literature. These issues should be kept in mind when interpreting the results of ozone health effects studies and, thus, when setting air quality standards.

Evidence for health effects of ozone

Long-term studies of ozone and mortality

Although most epidemiological studies of air pollution are time series, a few mortality studies have followed cohorts of individuals in an attempt to investigate the association between air pollution and mortality (Table 1). Although these studies have been called cohort studies, they really adopt a design that is a hybrid of cohort and ecologic designs. In these studies, cohorts of individuals from different cities are followed in time and deaths in the cohorts are recorded.

Table 1. Major long-term studies considering ozone*.

Reference	Location, period	Design, methods	Ozone concentrations	Results reported	Remarks
Krewski et al., 2000	151 metropolitan areas in the US. Individual subjects followed up for variable periods of time. Extension of the Harvard six cities and ACS II studies.	'Hybrid' design, i.e., some covariates were known on individual level, others, in particular air pollution, measured on the city level. Thus study is hybrid of cohort and ecologic designs. Cox proportional hazards and extensions used for analyses.	Average daily 1-h concentrations of ozone used in analyses. Levels not reported.	No associations with average daily 1-h concentrations of ozone and end points examined (total, cardiopulmonary, and lung cancer deaths) in either of the 6-month periods (April–September and October–March).	Ozone data available for 117 of 151 metropolitan areas. Focus of the study was fine particles and sulfates. Analyses using peak ozone part of sensitivity analyses of PM effects.
Lipfert et al., 2000	National cohort of approximately 70,000 US veterans diagnosed with hypertension in the 1970s. About 21 years of follow-up.	Hybrid design in the sense defined above. Cox proportional hazards used for analyses. Peak ozone in four distinct exposure periods considered for analyses: pre-1974, 1975–1981, 1982–1988, 1989–1996.	The mean 95th percentiles for ozone in the four periods were as follows:1960–1974, 132 ppb;1975–1981, 140 ppb;1982–1988, 94 ppb;1989–1996, 85 ppb.	Among the pollutants NO_2 and peak ozone were associated with concurrent mortality risk, but only the former with delayed risk. The mortality excess relative risk associated with ozone was reported to be about 10% at an 'adjusted' mean. How this adjustment was made is not entirely clear. There was a suggestion of a threshold for ozone at about 140 ppb.	This study is noteworthy because the adjustment of selected ecological covariates was done at the zip code level. This procedure presumably leads to better adjustment than one done at the city level. Hypertensive veterans are almost certainly not representative of the general population. Hence the generalizability of results to other populations is questionable.
Lipfert et al., 2006a, 2006b	Update of the Lipfert et al., 2000 study above with follow-up through 2001.	Hybrid design. Cox proportional hazards model used for analyses. For measures of ozone, see above.	See above. For the period 1997–2001, the peak ozone concentration is reported to be 84 ppb.	Ozone is significantly associated with mortality in the period 1989–1996, but not in 1997–2001. The authors conclude that traffic ensity is a much better predictor of mortality than any component of the air pollution mix.	This study is noteworthy for using a precisely defined measure of traffic density rather than proximity to a major highway.
Jerrett et al., 2005	Extension of the ACS II study to Los Angeles. Twenty-three thousand subjects with about 5000 deaths over the period 1982–2000. Focus was on fine PM. Ozone exposures interpolated from 42 fixed-site monitors.	Hybrid design. Cox proportional hazards model used for analyses.	Not reported	No association of ozone with mortality with two measures of exposure, expected peak daily concentration and average of four highest 8-h maxima.	Ozone exposure interpolated from 42 fixed-site monitors to the zip code level.

*Since this Report was prepared, a paper by Jerrett et al. (2009) has been published that reports analyses indicating that long-term exposure to ozone may increase mortality from respiratory diseases.

Individual-level information on some covariates likely to confound the association between air pollution and mortality, such as cigarette smoking, is collected and used in the statistical analyses. However, information on exposure to air pollution is available only at the population level from central monitors in the cities from which the cohorts are drawn. Thus, inferences regarding the association between air pollution and mortality are based upon differences in the levels of air pollution in the different areas in which the study is conducted. Most of the long-term studies have focused on the association of fine PM and mortality. Some studies considered the gaseous pollutants, but only as possible confounders of the association of fine PM with mortality. Table 1 describes the main features of the major long-term studies that considered ozone.

Before discussing the studies that have examined the relationship between long-term exposure to ozone and mortality, it is appropriate to discuss briefly a recent paper by Janes et al. (2007) on the issue of unmeasured confounding in epidemiologic studies of fine PM and mortality. The paper

raises serious questions regarding the validity of previously reported associations between fine PM and mortality and, in turn, has important implications for the association between ozone and mortality. Janes et al. (2007) use a new approach to investigating confounding in air pollution studies. There have been substantial decreases in air pollution in the United States, including fine PM pollution, in parallel with decreases in death rates. It is difficult, however, to attribute the decline in death rates to a decline in pollution because of the myriad other changes in demographics and lifestyle that have also occurred over the same period of time.

The authors note that the association between national trends in fine PM and mortality "is likely to be confounded by slowly time-varying factors, such as changes in industrial activities and the economy, improving health care, and large scale weather events." However, these associations at the local level are less subject to confounding and, therefore, a positive association detected at this level would be more likely to reflect a causal association between fine PM and mortality. Moreover, if fine PM pollution is causally associated with mortality, then areas of the country that have seen large declines in fine PM pollution should also see larger declines in mortality than areas of the country in which there have been more modest declines in fine PM pollution.

To test the hypothesis that declines in fine PM pollution are causally associated with declines in mortality, they use a statistical approach that decomposes the association between fine PM and mortality into a contribution at the national level and another at the local level. They analyze the association between fine PM and mortality in 113 U.S. counties over the 3-year period 2000–2002, and report associations between fine PM and mortality at the national, but not the local, level, and conclude that "if the association at the national scale is set aside, there is little evidence of an association between 12-month exposure to $PM_{2.5}$ and mortality." This conclusion suggests that the reported associations between fine PM and mortality are not causal but can be explained by confounding. Strengths of the study include the number of counties included in the analyses, the robustness of results to sensitivity analyses, and the use of regression calibration methods to adjust for possible measurement error. In view of these results and given the strong associations of PM and mortality reported in other studies, one can only wonder what would happen if similar analyses were conducted with ozone.

The long-term studies of air pollution and mortality, which were central to the Agency's decision on the NAAQS for PM, play little or no role in the Agency position on ozone. The Staff Paper presents a brief discussion of long-term studies of air pollution and mortality. With the exception of the Veterans' study discussed below, these studies reported no evidence of an association between ozone concentrations and mortality. As Krewski et al. (2000) note in their reanalyses of the Harvard Six Cities Study, "The Six Cities Study, with its small number of cities and high degree of correlation among the air pollutants monitored, did not permit a clear distinction among the effects of gaseous and fine particle pollutants. Indeed, estimates of the relative risk of mortality from all causes were similar for exposure to fine particles, sulfate, sulfur dioxide, and nitrogen dioxide. Of the gaseous co-pollutants in the Six Cities Study, *only ozone did not display an association with mortality.*" Similarly, Krewski et al. (2000) found little evidence of association between ozone and mortality in their reanalyses of the ACS II study. In a AQ4 study in Los Angeles based on the A[illegible] cohort, Jerrett et al. (2005) found no evidence of an association between ozone and mortality.

The U.S. Veterans' Cohort Study (Lipfert et al, 2000) is the only long-term study of air pollution and mortality to report significant associations between ozone and mortality. The cohort consists of approximately 70,000 U.S. veterans who were diagnosed with hypertension in the mid 1970s. The cohort had an average age of about 51 at recruitment, is all male, and is about 65% white and 35% non-white. In addition to air pollution variables based on county of residence, which were considered in some detail, information on individual level covariates, such as smoking, were included in the analyses. In contrast to the original Six Cities and ACS II studies, all measured criteria pollutants, with the exception of lead, were considered in the analyses. As in the Harvard Six Cities and the ACS II studies, the basic analytic tool was Cox proportional hazards regression. Four different exposure and three different mortality periods were considered, yielding a total of 12 distinct exposure and mortality period combinations for each pollutant. Among the pollutants, the strongest associations were seen with NO_2 and peak ozone. Of these two pollutants, the authors reported that ozone showed the stronger association with mortality, although there was an indication of a threshold at about 0.14 ppm for ozone effects. No significant PM association was seen with any of the various measures used (total suspended particulate [TSP], PM_{10}, sulfates, fine PM). The authors point out, however, "it must be recognized that all potentially harmful pollutant species are not measured routinely and thus cannot be included in epidemiology studies of this type. For this reason, those pollutants that are included should be considered as indices of the overall urban pollution mix. Further the nature of this mix has changed significantly during the period evaluated in this study."

Whereas the original Veterans' Cohort Study (Lipfert et al., 2000) was briefly discussed in the Staff Paper, a recent update was not (Lipfert et al., 2006a, 2006b). The updated study extends the mortality follow-up of the Veterans' Cohort through 2001 and considers data on county-level traffic density as a predictor in the regression analyses. The authors report that traffic density is a better predictor of mortality than any of the ambient air quality measures, including fine PM:

> *"Traffic density is seen to be a significant and robust predictor of survival in this cohort, more so than ambient air quality, with the possible exception of ozone. Stronger effects of traffic density are seen in the counties that have ambient air quality monitoring data, which also tend to*

have higher levels of traffic density. These proportional-hazard modeling results indicate only modest changes in traffic-related mortality risks over time, from 1976–2001, despite the decline in regulated tailpipe emissions per vehicle since the mid-1970s. This suggests that other environmental effects may be involved, such as particles from brake, tire, and road wear, traffic noise, psychological stress, and spatial gradients in socioeconomic status."

Lipfert et al. (2000, 2006a) reported associations between peak ozone, but not mean ozone, and mortality in the original and updated analyses. As noted above, the authors reported evidence of a threshold for peak ozone at about 0.14 ppm based on analyses of deaths during the period 1982–1988. The association appeared to be strongest for mortality in the period 1989–1996, but apparently the authors did not address the issue of a possible threshold. For the period 1997–2001, the association was not statistically significant.

The EPA Staff Paper (2007a) concluded that "consistent associations have not been reported between long-term ozone exposure and all-cause, cardiopulmonary or lung cancer mortality." The results of the long-term studies raise issues in the interpretation of time-series studies of ozone and mortality. Kunzli et al. (2001) distinguish four possibilities regarding the association of air pollution with mortality: "1) air pollution increases both the risk of underlying diseases leading to frailty and the short-term risk of death among the frail; 2) air pollution increases the risk of chronic diseases leading to frailty but is unrelated to timing of death; 3) air pollution is unrelated to risk of chronic diseases but short-term exposure increases mortality among persons who are frail; and 4) neither underlying chronic disease nor the event of death is related to air pollution exposure." They go on to argue that time-series studies capture deaths from categories 2 and 3, whereas long-term studies capture all the deaths associated with air pollution. They conclude that time-series studies underestimate the number of deaths attributable to air pollution and recommend that estimation of the impact of air pollution on mortality be based on long-term studies. In a later paper, Burnett et al. (2003) offer an alternative approach to understanding the relationship between time-series and cohort studies. They partition the hazard function into components related to long-term and short-term exposures to air pollution. This approach provides a framework for understanding the discrepant findings of the time-series and long-term studies of ozone and mortality. At the very least, the Agency should address the conundrum raised by the absence of an ozone effect in long-term studies. As one of the reviewers of this report suggested, perhaps the underlying mechanisms for producing effects are different.

Short-term morbidity and mortality (time-series) studies

In a time-series study, daily data are collected over a number of years, covering mortality or some other end point of public health interest (e.g., hospital admissions, asthma attacks in children). These responses are then included in a Poisson regression analysis where the explanatory variables include long-term trend and/or seasonality, meteorology, and air pollution. Long-term trend and seasonality are typically modeled as a smooth function of time, either through some expansion (e.g., splines) or through the generalized additive modeling (GAM) approach. Meteorology is modeled through a variety of different approaches, most often involving temperature and dew point. Finally, the air pollution variable of interest (in this discussion, ozone) can be modeled either linearly or nonlinearly, using a variety of lags, and either on its own or in conjunction with other pollutants (co-pollutants). In ozone studies, the most commonly used lag is 0 (in other words, current day's ozone is used as a predictor of current day's mortality), but it is also common to study ozone at a lag of 1 or 2 days, or some average over lags 0 and 1 or lags 1 and 2. An alternative model is the distributed lag model, in which ozone lags of up to 6 days are included in the model with separate regression coefficients for each lag. Day of week is also typically included as an explanatory variable, and some studies have been restricted to summer months because it was found that summer data produce the most significant association for ozone. All of this analysis is done initially one city at a time, but in recent years many studies have been multi-city. The inspiration for multi-city approaches was the NMMAPS study, which started as a study of particulate matter (PM) effects but in recent years has been extended to include ozone. NMMAPS stands for the National Morbidity, Mortality and Air Pollution Study, based at Johns Hopkins University.

Multiple city results

The main results of the NMMAPS study on ozone are summarized in the paper by Bell et al. (2004). This study used data from 95 U.S. cities covering the period 1987–2000. The authors used a distributed lag model with data on 24-h ozone averages for lags 0–6. Long-term trends were modeled through smoothing splines of between 7 and 21 degrees of freedom per year. Meteorology was modeled through temperature at lag 0, the average of temperature at lags 1–3, dew point at lag 0, and the average of dew point at lags 1–3, each modeled nonlinearly through smoothing splines. Interaction terms were included to separate mortality counts into three age groups (<65, 65–74, 75 and over) through a common regression function applied for all three groups. The analyses covered year-round data though many cities were implicitly restricted to summer months because only summer ozone data were available. After calculating a single regression coefficient, with corresponding standard error, to represent the overall change in ozone associated with a 10-ppb rise in 24-h average ozone in each city, the results were combined across cities using a Bayesian hierarchical analysis. Although one set of results was expressed at the level of individual cities, most of them were summarized as an overall national average relative risk.

The result of this analysis was that a 10-ppb rise in 24-h average ozone was associated with a 0.52% rise on total mortality (excluding accidental deaths), with a 95% Bayesian credible interval (equivalent for all practical purposes to a

95% confidence interval) of 0.27%–0.77%. This is for a distributed lag model: the corresponding result for lag 0 alone was 0.25%, credible interval 0.12%–0.39%. If deaths were restricted to cardiovascular and respiratory causes, the corresponding relative risk estimate based on a distributed lag model was 0.64%, credible interval 0.31%–0.98%. Using the same model to analyze the three age groups separately resulted in slightly different estimates, e.g., the 65–74 age group has an estimated relative risk of 0.70% (credible interval, 0.28%–1.12%), with results for the other two age groups almost identical to those from the combined age group model. If results were restricted to summer days (April to October) the estimated relative risk actually went down, from 0.52% to 0.39% (credible interval, 0.13%– 0.65%).

The question of a co-pollutant effect due to PM is complicated by the fact that most cities do not have daily records of either PM_{10} or $PM_{2.5}$. Therefore, comparisons are only possible using single-day lags. The paper included single-city comparisons of the estimates computed both with and without a PM_{10} adjustment, though no overall national estimate was quoted for the relative risk estimate with PM_{10} adjustment. Nevertheless it was claimed that the overall results were robust to the inclusion of either PM_{10} or $PM_{2.5}$. This result has been challenged by Smith et al. (2009) who showed that when PM_{10} is included as a co-pollutant, the estimated ozone effect typically goes down by about 25%, which they view as being a significant effect.

Another recent paper on PM confounding for ozone is Bell et al. (2007). This paper extended the results of Bell et al. (2007) by including numerous other analyses along similar lines to the earlier paper. The overall conclusion that the authors reached was still that confounding by particulate matter is not an issue in ozone studies. However, Table 2 of that paper does show about a 25% reduction in the point estimate of the ozone effect, when PM_{10} is included as a confounder. Thus, there still seem to be differences of interpretation of these analyses.

Other multi-city studies were also covered in the EPA Staff Paper (2007a). In another paper from the NMMAPS group, Huang et al. (2005) analyzed data for cardiovascular and respiratory mortality for 19 cities from 1987 to 1994 (compared with 95 cities, 1987–2000 for the full NMMAPS study). The data were restricted to summer months (June through September). The analytical method was essentially the same as in Bell et al. (2004) though is different in some details, e.g., the age group effect was modeled only as a single indicator variable for each age group rather than smooth curves as in Bell et al. (2004). For the possible confounding effect of PM_{10}, only single-lag models were appropriate for the reason discussed earlier: for both ozone and PM_{10} at lag 2 (the lag where the discrepancy between the with-PM_{10} and without-PM_{10} results was largest) the estimated ozone effect was 0.64% (credible interval, 0.17–1.07%) when PM_{10} was not included in the model, and 0.46% (–0.32%–1.17%) when PM_{10} was included. This is consistent with several other results reported later, that suggest the ozone effect is generally attenuated when PM_{10} is also included in the model. In this case, some confounding was also noted with either NO_2 or SO_2 included (singly) as co-pollutants.

Schwartz (2005) also performed a multi-city study, using 14 U.S. cities, but using the "case-crossover" design instead of a time-series analysis. In this design, the date on which an individual died is matched against an alternative date on which the individual did not die, and the ozone readings for the 2 days compared. To minimize the effect due to seasonality or long-term trend, the matching date is held close to the death date. Schwartz (2005) also matched for temperature. He found that a 10-ppb rise in ozone was associated with a 0.57% (95% credible interval, 0.02%–1.1%) increase in deaths, for the full-year analysis. It should be noted that Schwartz (2005) actually used daily 1-h maximum ozone, not 24-h average as in the NMMAPS papers, but for the results reported here, in order to make comparisons with results from other papers, we have converted to 24-h average ozone using the conversion factor that a 1-ppb rise in 24-h ozone corresponds to a 2.5-ppb rise in 1-h daily maximum ozone (Thurston and Ito, 2001). Later in this report we will discuss the issue of using conversion factors for converting concentration-response coefficients from one ozone metric such as 1-h maximum concentration to a second metric such as 24-h concentration. With that conversion, the main result is very similar to the main result of Bell et al. (2004) (the credible interval is wider, but that is most likely a consequence of the smaller number of cities used in the analysis, not to mention an alternative study design). Schwartz (2005) also reported, however, that when restricted to winter months, no association was found (point estimate –0.32%, credible interval –1.32%–0.7%) but when restricted to summer months, an increased relative risk was estimated (0.92%, credible interval 0.27%–1.55%). In a slightly different analysis in which temperature was accounted for through regression rather than by matching, Schwartz (2005) found that including PM_{10} as a co-pollutant made no difference at all to the point estimate (0.47% in each case).

As part of a meta-analysis to be discussed in more detail later, Ito et al. (2005) carried out a time-series analysis on seven U.S. cities in order to look at model sensitivity issues. They analyzed the data by methods similar to Bell et al. (2004, 2005) but using four different meteorological models, from a "quintiles indicator variables" approach that they attribute to Moolgavkar et al. (1995), to a "4 smoother" model that is effectively equivalent to the meteorological adjustment approach of Bell et al. (2004, 2005). Relative risk estimates were computed separately for winter and summer data though the winter estimates were generally not statistically significant. Comparisons were also made for ozone relative risk estimates with and without either PM_{10} or $PM_{2.5}$ as a co-pollutant. The results were combined across six of the cities (the seventh city, New York, was not included because no PM data were available). The biggest difference among the results was associated with the different weather adjustment models. For example, for all-year data without PM the combined estimate was a rise of 1.0% (0.55%–1.45%) associated with a 10-ppb rise in ozone based on the quintiles approach

to meteorology, versus 0.5% (0%–1.0%) using the 4 smoother approach. For summer results including a PM adjustment, the corresponding results were 1.0% (0.3%–1.7%) using the quintiles approach and 0.55% (–0.05%–1.1%) using the 4 smoother approach.

The final multi-city analysis we discuss here is Gryparis et al. (2004) based on 23 European cities in the APHEA project (Air Pollution and Health: A European Approach). We note in passing that although this is not a U.S. study, it was cited in the EPA Criteria Document (2006) and some of the meta-analyses that are reported later also combined results from U.S. and non-U.S. cities. For these reasons, we believe it is important to note when results from U.S. and non-U.S. cities produce substantially different estimates. Gryparis et al. (2004) used a very similar approach to Bell et al. (2004), though they did not describe the long-term trend and meteorology corrections in enough detail to make a precise comparison. In their study, ozone is measured in $\mu g/m^3$, which we convert to ppb at the rate of 1.96 $\mu g/ m^3$ equals 1 ppb (Bell et al., 2005). They also used 1-h daily maximum ozone, which we convert to 24-h average ozone using the same 2.5 conversion factor noted previously. With these changes, their main result is that a 10-ppb rise in 24-h ozone is associated with a 1.62% rise in total mortality, with a 95% credible interval of 0.83%–2.55%.

Possible explanations for the different estimates include different exposure patterns in European cities (due, e.g., to greater use of public transport), different placement of monitors in European cities, and differences in the statistical modeling strategy. It is also possible that ozone is serving as a proxy variable in both the United States and European studies for some missing causative factor and that the ratio between levels of ozone and the causal variable might be different for the two continents as well as within each continent.

The published multi-city studies put a huge emphasis on the overall nationally averaged effect, but this ignores the very real differences that are evident among different cities. For example, in the NMMAPS analysis, our Figure 2 essentially reproduces Figure 2 from Bell et al. (2004), showing the posterior estimates (calculated by the Bayesian hierarchical approach) of ozone-mortality coefficients in each of the 98 cities in the study. Figure 2 in this report extends the Bell et al. (2004) figure, independently recalculated by Smith et al. (2009). In addition to reproducing the posterior estimates of Bell et al. (2004), this figure also shows the raw estimates and posterior estimates under an alternative "regional prior." The substantial variability among the raw estimates for the 98 cities is very evident. The posterior estimates under the regional prior also differ from those under the national prior. The analyses by both Bell et al. (2004) and Smith et al. (2009) yielded statistically significant coefficients for only a few cities (New York, Newark, Philadelphia, Dallas-Fort Worth, Chicago, and Houston), with the vast majority of the cities not having statistically significant coefficients. Moreover, the raw estimates shown in Figure 2 (before the Bayesian part of the calculation) show a far greater variability among the cities, with about a third of the coefficients negative. In our view, it is very important to take into account the variability of the ozone-mortality coefficients in the individual cities, especially when the coefficients are applied to risk analyses, but this does not appear to have been done in current EPA analyses.

In Figure 3, taken from Smith et al. (2009), regional weighted averages of the posterior mean prior estimates are shown. It is apparent that statistically significant positive coefficients were obtained only for the Northeast, Industrial Midwest and Southeast. In other areas of the country, there is no apparent effect of ozone on mortality.

A national coefficient is shown in Figure 2 following the example of Bell et al. (2004). A national coefficient is also shown in Figure 3. In our view, the national coefficients are of limited value in view of the substantial variability that is evident on a city-to-city basis and from one part of the country to another.

Another issue raised by the NMMAPS papers is that with one exception, they calculated all their ozone-mortality coefficients based on 24-h ozone. In our view, they should have used the maximum 8-h ozone in any day, because the EPA standard is based on this variable. As an example, Figure 4 shows estimated coefficients for individual cities and Figure 5 shows the corresponding regional values based on the maximum 8-h ozone metric. Once again we note that only a handful of the cities have statistically significant ozone-mortality coefficients, whether these are assessed by the raw or posterior estimates. The one exception to citation of an 8-h ozone metric is that Bell et al. (2004) did calculate a national average effect for the 8-h average. The value obtained is consistent with the national effect coefficient shown in Figures 4 and 5.

As noted earlier, a number of investigators have converted ozone concentration–health effects coefficients obtained with one metric, such as the 1-h maximum concentration, to a second metric, such as the 24-h average concentration. The validity of this approach is open to question, as demonstrated by Smith et al. (2009) who calculated concentration-mortality coefficients for each of the 98 cities using all three time metrics. Even for a single city, it is not well established that the concentration-mortality coefficient for the three metrics (1-h maximum, 8-h maximum, and 24-h average) have a consistent relationship. It is quite possible that one of the three metrics may be more strongly associated with a particular health outcome, such as excess mortality, than the other metrics. The relationship between the three metrics varies over time for any given city and is certainly variable among cities. The association between concentration and excess mortality across the cities in the NMMAPS data set for all three metrics is quite variable, as demonstrated by Smith et al. (2009). Scatter plots of the relationship between the 24-h and 8-h posterior estimates and between the 1-h and 8-h posterior estimates are shown in Figure 6. It is readily apparent that the relationship of the coefficients for the different metrics for individual cities is so variable that use of a single "national" conversion coefficient for multiple cities

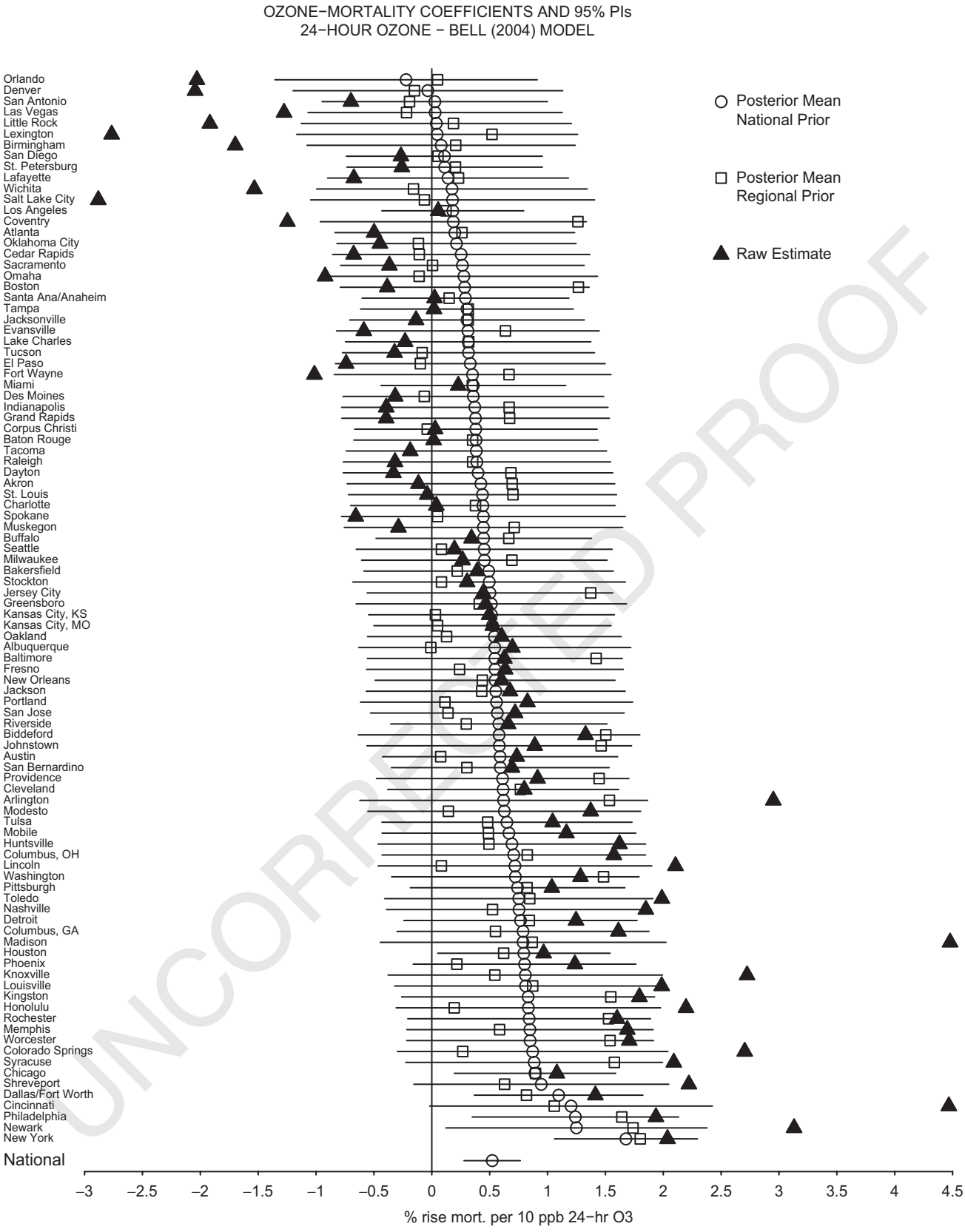

Figure 2. Ninety-five percent posterior intervals for the ozone-mortality coefficients, all-year data, by the hierarchical Bayesian method as in Figure 2 of Bell et al. (2004). The Bayesian posterior estimates under the "national prior" (circles) are shown alongside those for the "regional prior" (squares) and the raw maximum likelihood estimates (triangles) (Figure 1 of Smith et al., 2009).

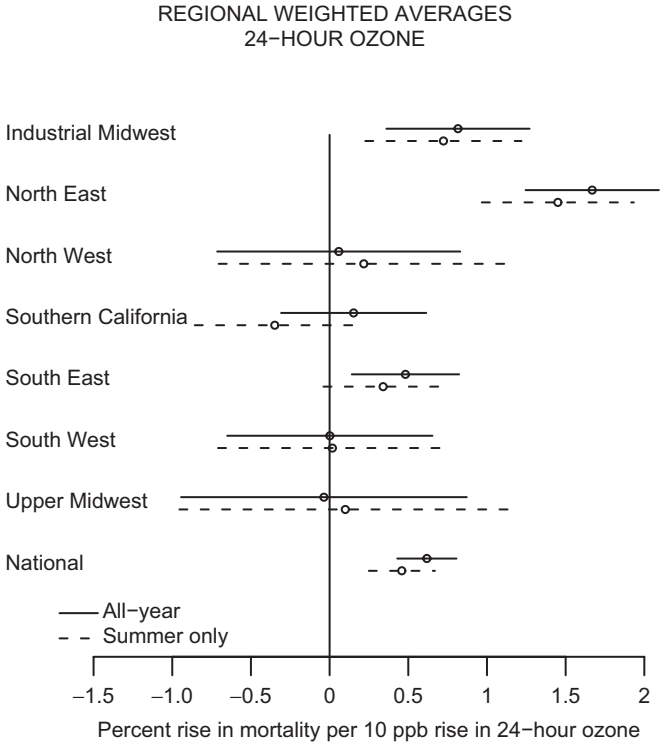

Figure 3. Regional estimates of population-weighted average regression coefficients based on 24-h ozone, with 95% PIs (based on data in Table 3 of Smith et al., 2009).

in a meta-analysis would lead to erroneous estimates of the converted coefficient for many cities.

Meta-analyses and evidence on publication and model selection biases

Meta-analyses are different from multi-city time-series studies in that they rely on previously published studies; there is therefore the possibility of *publication bias* owing to the tendency that statistically insignificant results may not get published at all. A second distinction between multi-city analyses and meta-analyses is that in the former, it is possible to ensure that exactly the same regression model is fitted in each city, whereas in meta-analyses this is hard to control, raising the possibility of *model-selection bias*, i.e., the bias that results when many models are compared but only the largest or the most statistically significant association is reported. A recent collection of papers in the journal *Epidemiology* provides evidence of both kinds of bias.

Bell et al. (2005) analyzed data from both U.S. and non-U.S. cities. The studies used different lags from 0 to 2, but Bell et al. (2005) used results from lag 0 where possible to maintain the greatest comparability among studies (they also showed that the lag 0 studies led to the largest estimates overall). None of the studies in the meta-analysis used distributed lag models. Data were combined across different studies using Bayesian hierarchical models, similar to their multicities study (Bell et al., 2004). This method is theoretically superior to the non-Bayesian techniques used in other meta-analyses because it takes into account the uncertainty in estimating the intercity variance, and thereby leads to slightly wider credible intervals, though for most practical comparisons the Bayesian and non-Bayesian meta-analysis methods should lead to similar results. Where different types of mortality were concerned, the estimates for cardiovascular mortality were higher than those for total mortality, but those for respiratory mortality were lower. They looked at whether estimates were changed by including a PM adjustment, concluding that they were not. Table 3 of their paper showed much wider credible intervals in the case of PM-adjusted estimates, but this may be because only a subset of the studies used a PM adjustment. In comparisons between year-round and summer-only estimates, they found that summer-only studies produced almost twice the estimated relative risk (1.50% with a credible interval from 0.72% to 2.29%) for total mortality based on combined U.S. and non-U.S. data. The relative risk estimated for all-season total mortality was 0.87% (credible interval, 0.55%–1.18%).

The meta-analyses reported by Bell et al. (2005) were kept separate from the NMMAPS results of Bell et al. (2005). However, eight of the meta-analysis cities were also NMMAPS cities. For lag 0 ozone, all-year total mortality, the estimate of Bell et al. (2004) was 0.25% (credible interval 0.12%–0.39%) for 95 cities in NMMAPS. This is to be compared with the just-quoted meta-analysis result of 0.87%, which seems clear evidence of a publication bias in the use of previously published results. However for the eight cities that are common to both studies, Figure 2 in the paper by Bell et al. (2005) also shows the individual-city estimates and confidence intervals. Strikingly, in every case the estimate included in the meta-analysis is larger than the NMMAPS estimate for that city. Bell et al. (2005) attribute this result to publication bias but it seems unlikely that a discrepancy could arise for this reason alone (e.g., even if these eight cities had been selected from a larger set of cities as the ones with the largest ozone effects, that is not sufficient reason why the estimates should be different when reanalyzed by NMMAPS). At least part of the reason must be the difference in modeling approaches—the fact that NMMAPS was careful to use the same model in all cities, whereas the other published papers, by different authors using different modeling strategies, were more likely to have selected the model to maximize the relative risk estimate. In other words, these comparisons show evidence of model selection bias as well as publication bias.

Ito et al. (2005) used results from 11 U.S. and 25 non-U.S. cities, with a "selected" lag of up to 3 days in each city (they did not say how they made the selection). They looked at mortality effects by season and confounding by PM. Results from different cities were combined by a non-Bayesian meta-analysis procedure (DerSimonian-Laird). Overall they found a risk increase of 0.8% (95% confidence interval, 0.55%–1.0%) associated with a 10-ppb rise in 24-h ozone. However, among studies reporting a seasonal breakdown, the results were 1.1% (0.4%–1.8%) for the all-year relative risk and 1.75% (1.05%–2.45%) for summer, again showing clearly that relative risk estimates are higher during the summer. The results showed little effect due to confounding by

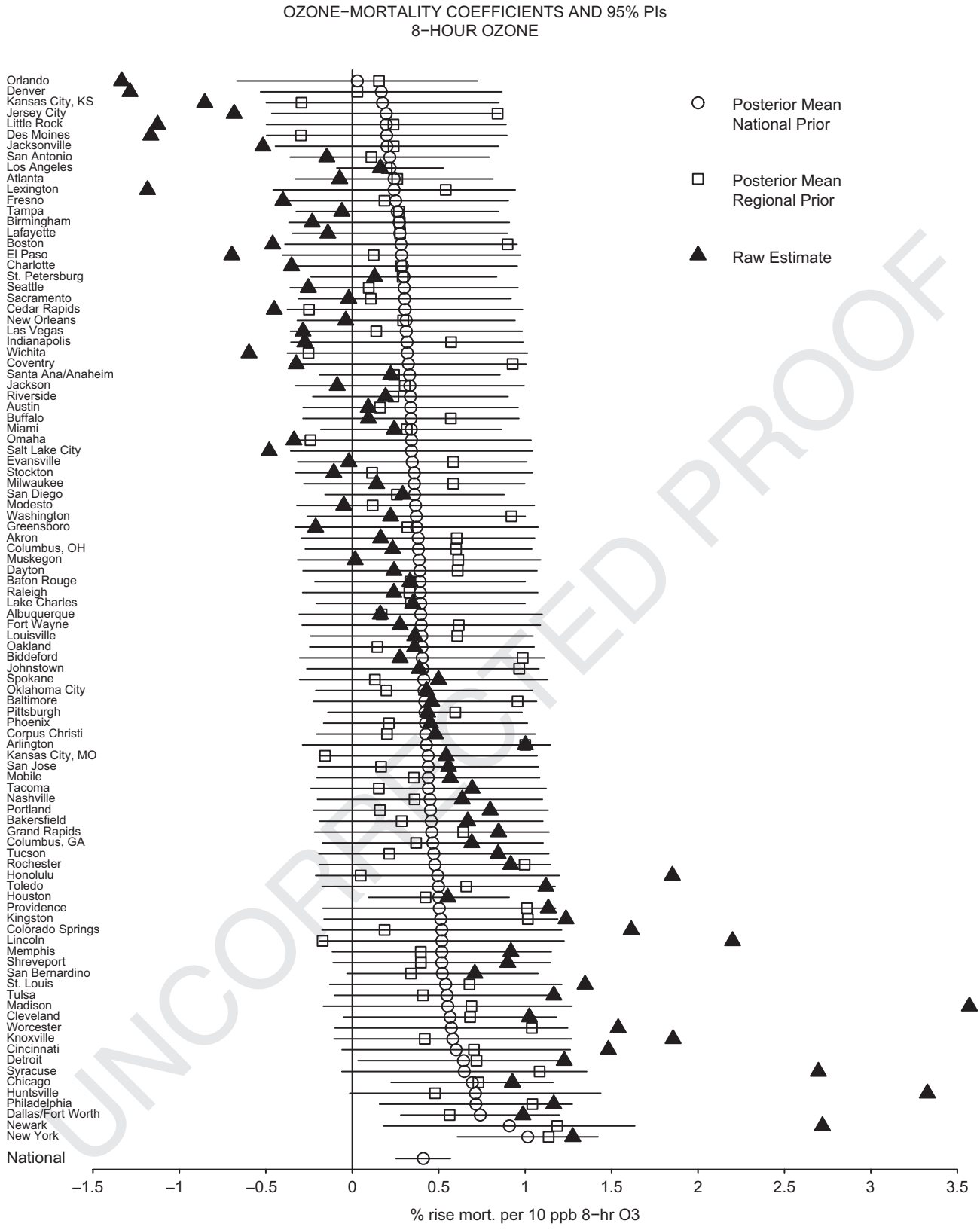

Figure 4. Ninety-five percent posterior intervals for the ozone-mortality coefficients, based on 8-hour ozone, all-year data. The Bayesian posterior estimates under the "national prior" (circles) are shown alongside those for the "regional prior" (squares) and the raw maximum likelihood estimates (triangles) (Figure 4 of Smith et al., 2009).

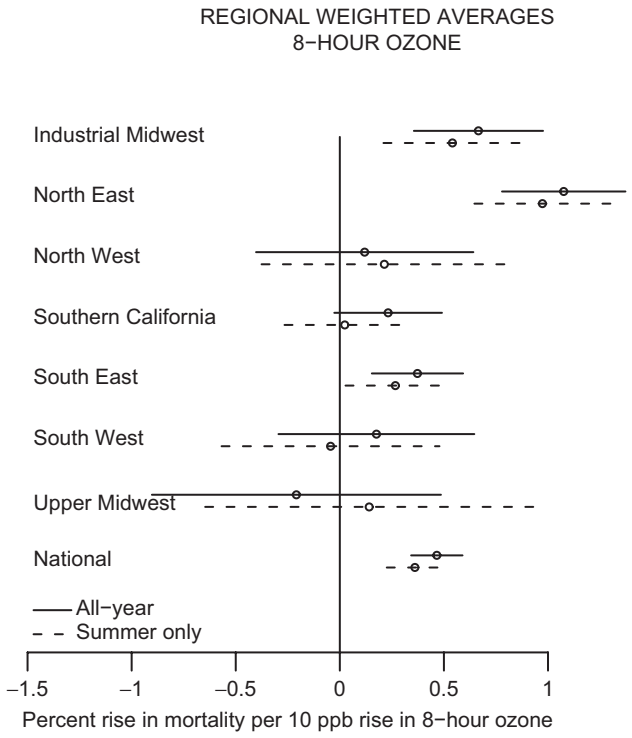

Figure 5. Regional estimates of population-weighted average regression coeficients based on 8-h ozone with 95% PIs (based on data in Table 3 of Smith et al., 2009).

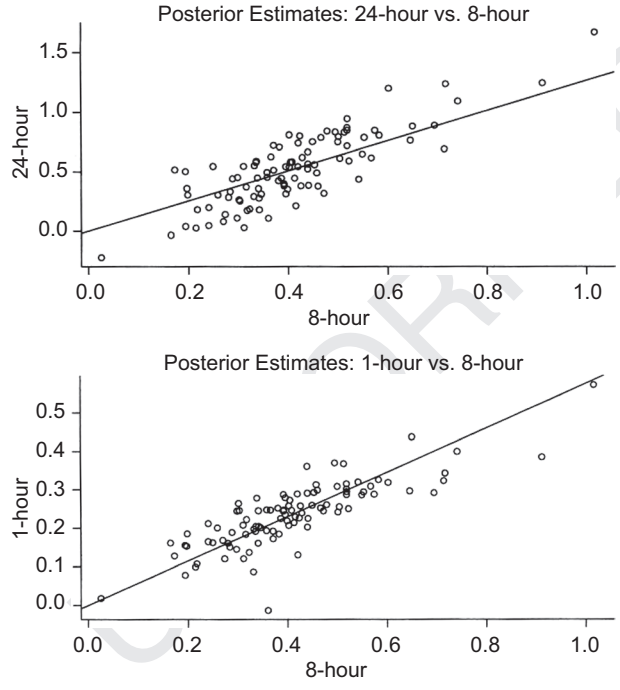

Figure 6. Scatter plots of posterior estimates corresponding to 24-, 8-, 1-h ozone (Figure 3 of Smith et al., 2009).

PM, though in cases where both sets of relative risks were computed (with and without PM), the estimates with PM were lower. There was some evidence of publication bias as determined through an asymmetry test. In some cases, different studies using the same city's data showed marked contrast in the results, which is further evidence of a model selection bias. As noted already, in the same study, Ito et al. (2005) also did a multi-city analysis with seven U.S. cities, which amplified the issue of model-selection bias.

The third meta-analysis, by Levy et al. (2005), found a combined risk increase of 0.21% (confidence interval, 0.16%–0.26%) associated with ~~a 10 μg/m^3 in 1 h daily~~ maximum ozone. Using the same conversion factors defined earlier, this translates to an estimated relative risk of 1.03% (0.78%–1.27%) associated with a 10-ppb rise in 24-h average ozone. Thus the magnitude of the association was quite similar to those of the other two meta-analyses (but still substantially higher than NMMAPS). An original feature of their analysis was that it accounted for air-conditioning use (using publicly available Census data), showing that the ozone-mortality association is higher where there is lower use of central air conditioning. Some subsequent papers such as Bell and Dominici (2008) and Smith et al. (2009) have shown similar effects in the NMMAPS data. The implication of these analyses is that people who have central air conditioning in their homes are exposed to much lower ozone when indoors in summer.

Taken together, the meta-analyses provide evidence of publication bias and model selection bias. Particularly striking are the contrasts between the three fairly similar relative risk estimates from the meta-analyses, and the much lower estimate in NMMAPS. These analyses also provide evidence that model-selection bias may alter the magnitude of the estimated association by a factor of 2, or even 4, if the distinction between the NMMAPS and AHPEA results cannot be attributed to other differences in both air pollution and mortality patterns between United States and European cities.

Nonlinear exposure-response relationships

All the studies reported so far have used essentially a linear ambient concentration-response curve (on a log scale, i.e., the logarithm of expected deaths is linear in ozone). Bell et al. (2006) looked in the NMMAPS data for evidence both of a threshold (no change in mortality associated with ozone below a certain threshold) and for a nonlinear ambient concentration-response curve (in which the entire curve of expected deaths versus ozone is constructed as smoothly nonlinear). The evidence for a threshold, if there were one, would be of obvious relevance to the determination of a standard, but even in the absence of a threshold, the use of a nonlinear ambient concentration-response curve, if it could be reliably estimated, could potentially be of great value is risk assessment.

For this analysis, Bell et al. (2006) did not attempt a distributed lag model (which would involve even more technical complications), but confined themselves to the average 24-h ozone at lag 0 and lag 1 as the main ozone-related predictor of mortality. When the earlier linear concentration-response analysis was repeated with this measure of ozone, the estimated rise in mortality per 10 ppb increase in ozone was 0.32%, with a 95% credible interval from 0.17% to 0.46%.

This is compatible with the results of Bell et al. (2004), who showed a lower relative risk when confined to any single lag, but a higher relative risk using a distributed lag model. However, Bell et al. (2006) repeated the analysis on subsets of days defined by various thresholds—in one analysis, they fitted the same model but restricted to days on which the lag 0 and lag 1 average of 24-h ozone was less than *s*, where the values of *s* ranged from 60 ppb down to 5 ppb. For values of *s* down to 30 ppb, there was very little change in the estimated relative risk, which remained clearly statistically significant according to the authors' own Figure 2. Their own interpretation of this result was that it proved ozone still had deleterious health consequences at levels much lower than the current ozone standard.

We have several comments about this analysis. The first is that the conclusions in their paper are not entirely supported by the evidence presented—for example, they say "Daily changes in ambient O_3 were significantly associated with daily changes in the number of deaths…even when we used data that included only days with lag 01 average O_3 levels < 15 ppb." However, the result they present for 15 ppb is clearly not statistically significant.

A second comment is that there seems something odd about the claim for significant effects at very low levels of ozone—not only well below the current ozone standard but even the vast majority of the data. Observing the same effect in a more restricted study (for the city of Vancouver), Vedal et al. (2003) questioned whether such results could truly be indicative of a causal ozone effect, suggesting that the unmeasured effect of other air pollution or uncontrolled features of meteorology may be responsible.

However, another issue here concerns the relevance of concentrating on low ozone values anyway, given that the ozone levels of most relevance as far as the standard is concerned are between 60 and 80 ppb of 8-h ozone. Smith et al. (2009) estimated a piecewise-linear ozone-mortality association based on separate linear relationships for the ranges 0–40, 40–60, and 60–80 ppb (8-h ozone). They found, consistent with Bell et al. (2006), that the association on 0–40 ppb is statistically significant, as is that on 40–60 ppb, but they did not find a statistically significant association between 60–80 ppb except in the Industrial Midwest region. Only by combining data across the three ranges (in effect, assuming a linear ozone-mortality relationship across 0–80 ppb of 8-h ozone) were they able to establish a statistically significant association, but this raises the question of why do a nonlinear analysis at all. In summary, it looks as though Bell et al. (2006) looked at the wrong range by focusing on very low ozone exposures rather than concentrating their analysis on the range of interest for determining the standard.

The second part of the analysis by Bell et al. (2006) was aimed at estimating a nonlinear dose-response curve. The stated method of doing this was via a spline approach with knots at 0, 20, 40, and 80 ppb (but not 60 ppb). The result shown in Figure 3 of Bell et al. (2006) showed a steadily increasing ozone association above about 15 ppb, and confidence bands that indicate a statistically significant association above about 40 ppb. In our own analyses, we have not succeeded in reproducing this exact result, though we have shown similar results using a piecewise linear approach (Smith et al., 2009). Given these and other concerns, we believe the whole question of nonlinear dose-response curves is still very much undecided at the present time. Nevertheless, we recognize it as a very important topic of future research.

Exposure error

A number of studies have examined correlations between personal and ambient exposure for both gaseous pollution and PM. In general, they have found that correlations are higher for either $PM_{2.5}$ or particulate sulfate (SO_4^{2-}) than for ozone. For example, in a study using personal monitors for three groups of sensitive subpopulations in Baltimore and Boston, Koutrakis et al. (2005) noted that "ambient concentration of gaseous pollutants serve as a better surrogate for personal exposure to $PM_{2.5}$ than for personal exposure to gaseous pollutants," though with the qualification that the ozone results were hard to determine because many were below the detection thresholds. Sarnat et al. (2006) reached a similar conclusion in a study conducted in Steubenville, Ohio, which showed "strong associations between ambient particle concentrations and corresponding personal exposure," but "most associations between ambient gases and their corresponding exposures had low slopes and R^2 values" (though the associations were still statistically significant). The EPA's Staff Paper (2007a) (section 3.4.2.1, pages 3-39 through 3-42) noted that Sarnat et al. (2005) found a statistically significant positive correlation between personal and ambient exposures to ozone, but failed to cite Koutrakis et al. (2005) at all, or to note the implication that ambient ozone may in fact be acting as a proxy for PM in epidemiological studies.

The direct effects of measurement error on regression estimates from time-series mortality studies have been investigated for particulate matter but not for ozone. For example, Dominici et al. (2000) presented a Bayesian statistical analysis, though based on rather limited information, about differences between personal and ambient exposure to PM. Brauer et al. (2002) showed by simulation that exposure error could well result in failing to detect a health effects threshold in the case of $PM_{2.5}$, though not for SO_4^{2-} where the personal-ambient correlation is much higher. They did not do a corresponding simulation for ozone, but because the other studies just cited have shown that the personal to ambient correlation is even less for ozone than for $PM_{2.5}$, a similar simulation for ozone would most likely reach the same conclusion.

Conclusions for time-series mortality studies of ozone

Single-city studies of the association between ambient ozone and mortality show a very wide range of results, with both positive and negative regression coefficients that are generally not statistically significant, even when the analyses are for large cities and long time periods (e.g., 14 years in the

case of the NMMAPS study). However, a number of studies in recent years have shown a significant positive association in either multi-city studies (when time series from several cities are analyzed using the same statistical methodology and the results then combined) or in meta-analyses (in which previously published results are combined into a single overall estimate and confidence limit). Associations are strongest when confined to summer ozone; no study has found a significant association for winter data alone, though several (including the leading NMMAPS papers) have used all-year data without regard to season. All the studies employ corrections for long-term trend and seasonality, and include meteorological variables to avoid possible confounding. The ozone lags that are used vary considerably from one analysis to another, the most popular lags being days 0 (i.e., present-day ozone used as predictor of present-day mortality), 1, and 2. However, studies that have used a distributed lag model, in which the predictor of mortality is a weighted average over the past week, estimate the largest overall ozone relative risk.

Despite the impression given in EPA's criteria document and staff paper that these results are very consistent across different studies, closer scrutiny shows this not to be the case, even when the results are standardized with respect to ozone metric used (the results quoted here are for a 10-ppb rise in 24-h average ozone, to the extent we can make an exact conversion). The results in the three meta-analyses collected in Levy et al. (2005) all find ambient ozone-mortality coefficients more than three times larger than the corresponding NMMAPS estimates, which seems indicative both of publication bias (only the larger or more significant results being published) and of model selection bias (reporting only results that give the largest or most statistically significant regression coefficient). Although the NMMAPS authors (Bell et al., 2004, 2006) have taken the greatest care in both compiling a large dataset and trying to control for other effects, including weather, there remain doubts about whether all possible confounders have been correctly allowed for. Moreover, by quoting most of their results in terms of an overall national average, they have ignored the very real differences among cities that are apparent in close scrutiny of their own analyses. For example, there are strong ozone-mortality associations in Chicago, New York City, and a few other northeastern cities, and also Houston and Dallas, but other large cities have very slight or negligible associations, e.g., Los Angeles, Atlanta, and Miami. There are a number of western cities, such as Denver, Salt Lake City, and Albuquerque, that were in compliance with the 0.084 ozone standard set in 1997, but are likely to be nonattainment with the new ozone standard set at 0.075 ppm. There is no evidence in any of the analyses of a significant ozone-mortality association for these cities.

None of the recent studies on ozone-mortality associations have looked explicitly at measurement error. However this is clearly a concern, because personal-ambient exposure correlations are lower for ozone than for particulate matter, and one recent study concluded that ambient ozone may be a better predictor of personal exposure to $PM_{2.5}$ than of personal exposure to ozone. This raises the possibility that observed associations between ozone and mortality may in fact be acting as a proxy for a PM-mortality relationship. Another study showed by simulation that for $PM_{2.5}$, measurement error may be responsible for failure to detect a threshold. Although a corresponding study has not been conducted for ozone, given the even weaker associations between ambient and personal exposures of ozone than exist for $PM_{2.5}$, it seems very likely that the same result would be true for ozone.

The question of confounding by PM may also be looked at by including PM_{10} or $PM_{2.5}$ directly in the analysis as a co-pollutant along with ozone. Where this has been done, the results have generally indicated a slightly weaker though still significant association with ozone. However, data availability issues complicate the comparison. In particular, in the United States most cities have collected PM_{10} data only once every 6 days, so in NMMAPS, it was not possible to conduct an ozone + PM_{10} analysis that would be directly comparable with the results for ozone alone. Even less data are available for ambient $PM_{2.5}$. Moreover differences in the degree to which exposures to each pollutant are correlated with actual individual exposures could allow ozone to serve as a proxy for one of the other pollutants, even if the other pollutant were to be included in the regression model.

Questions related to thresholds and nonlinear ambient ozone-mortality relationships remain open at the present time. The recent study by Bell et al. (2006) reported a significant ozone-mortality association even when the analysis was confined to days with an ozone level below 15 ppb, though graphs in the same paper do not entirely support this conclusion. Whether an association found at such low ozone concentrations indicates a causal effect is questionable. The same paper also constructed a nonlinear dose-response curve showing an increasing ozone-mortality association above 15 ppb and a statistically significant association above 40 ppb (of 24-h average ozone). However, we have doubts about the methodology and feel that judgment should be reserved on this conclusion at the present time.

Time-series studies of ozone and hospital admissions

Parallel to the above studies about time-series analysis and mortality, there have been corresponding studies related to morbidity, by various measures such as hospital admissions for respiratory or cardiovascular diseases. Many of these studies were available at the time of the 1996 ozone review and there have been only a handful of more recent studies.

As an indication of some of the studies used in the 2006 risk assessment, Thurston et al. (1992) found significant associations between ozone and respiratory hospital admissions in three metropolitan areas (New York City, Albany, New York, and Buffalo, New York), for three summers (June–August, 1988–1990). This time period included the summer of 1988, which had extreme pollution levels. Schwartz et al. (1996) reviewed the methodology used in assessing pollution-morbidity relationships in time-series analysis (which raises more or less the same issues as do time-series mortality

studies) and analyzed an illustrative dataset for Cleveland, Ohio, finding a significant association that, translated to a 10-ppb rise in 24-h ozone, would indicate a relative risk of 1.04, with a 95% confidence interval from 1.01 to 1.08 (in other words, the result was marginally significant). Linn et al. (2000) studied the effect of ozone, CO, NO_2, and PM_{10} with cardiopulmonary hospital admissions in Los Angeles from 1992 to 1995, finding the strongest association due to CO, and ozone relative risk coefficients that were either negative or not statistically significant. Likewise, Ito (2003) reanalyzed a dataset of Lippmann et al. (2000) that studied the association of several pollutants with hospital admissions in Detroit. The original study of Lippmann et al. (2000) had been affected by the "GAM bug" (Dominici et al., 2002)—this was a problem with some of the statistical analyses due to the use of inappropriate default convergence criteria in the statistical package SPlus, but the error was detected and corrected in all studies published in or after 2003. The results displayed by Ito (2003) focus almost entirely on particulate matter, but the original results of Lippmann et al. (2000) show in almost all cases no statistically significant association with ozone; the one exception (among many endpoints studied) was for hospital admissions due to heart failure, where they did find a small just-statistically significant association.

The hospital admissions literature differs from the mortality literature in that most of the papers consider data only for a single city or a very small group of cities (e.g., three in Thurston et al.. 2000). The one moderately larger study by Burnett et al. (1997) did find significant ozone associations with hospitalization for respiratory causes in 16 Canadian cities, although their methodology was much weaker than that used in recent multi-city mortality studies. For example, they did not include a detailed meteorology model.

The EPA's staff paper (2007a) compiled these and a number of other results into a diagram showing the associations of ozone with different health outcomes (Figures 3–4, pages 3–56). However, it is clear from this figure that most of the individual results are not statistically significant, even though positive coefficients outnumber negative ones.

In conclusion, the relationship between ozone and hospital admissions has been much less intensively studied than the corresponding relationship with mortality, with only one (comparatively small) multi-city study and no meta-analyses. Few new studies have been published since the 1996 EPA review (1996a), and the two post-2000 studies they cite do not support the conclusion of a statistically significant ozone association. All of these studies, however, are subject to the same limitations and uncertainties described for the time-series mortality studies because they use the same types of data for estimating exposures to pollution, and use the same statistical estimation methods.

Challenges of identifying effects of correlated pollutants

The best approach to analyzing epidemiological data, when multiple correlated pollutants might impact human health, has been under considerable debate in the scientific community. The traditional statistical approach for multipollutant analyses has been criticized on the grounds that introducing a correlated pollutant might attenuate the estimate of the 'true' relative risk of the pollutant under study. However, not including confounders such as co-pollutants can overestimate the relative risk of the pollutant under study. As a result, the same events are counted more than once when the population impacts of multiple single pollutant analyses are estimated. Thus, for example, the sum total of the estimated deaths due to pollutant A and the estimated deaths due to pollutant B, both derived from single pollutant models, can overstate the impact of the two pollutants in the real-world situation in which both pollutants coexist.

Panel studies

Panel studies generally involve a large number of repeated measurements of various indicators of effect in a relatively small group of well-defined subjects. As discussed previously (see "Exposure Assessment"), panel studies are one of the few designs in which individual-level exposure assessment (e.g., from personal monitoring) is feasible. With this design, even without personal monitoring, it is possible to characterize exposures with a high degree of accuracy and precision by limiting the study area and using study specific monitoring data. For example, studies of Korrick et al. (1998), Brauer et al. (1996), Kinney et al. (1996), and Delfino et al. (1996) are similar to earlier studies of children attending summer camps. Further, the design has the potential to focus on subjects with specified time-activity patterns. Generally, the weaknesses of panel study designs are their ability to be generalized to a larger population and the use of intermediate endpoints, often including measures of effect that may have undetermined clinical relevance, for example those in the study of Kinney et al. (1996).

New panel studies are listed in Table 8A-2 of the 2006 Criteria Document. In general, there are two types of panel studies that have been considered. Several studies (Brauer et al., 1996; Korrick et al., 1998) involved well-defined exposures to ozone in exercising individuals and were conducted in settings where impacts of co-pollutants were minimized. For example, Brauer et al. (1996) measured the association between ozone and morning and evening lung function in a group of outdoor berry pickers. Subjects were outdoors during the entire daytime period. Ambient monitors were at or very near the study locations and personal monitoring assessed and evaluated exposure in a subset of the study subjects (mean difference = 2.5 ppb, $r = 0.64$). Korrick et al. (1998) measured lung function of hikers on Mount Washington and assessed their association with measurements of ozone at the mountain summit and base. Both studies measured particle mass and acidity although only Korrick et al. (1998) directly assessed the impact of co-pollutants finding an effect for both $PM_{2.5}$ and strong aerosol acidity.

Another group of studies longitudinally assessed lung function or other measures in more complex exposure situations. For example, Mortimer et al. (2000), and Gent et al. 2003), used ambient monitoring network data to assess exposures of study subjects dispersed over relatively large areas.

In the case of Mortimer et al. (2000), all ambient monitors in the county were averaged for each of eight urban study locations, in an approach similar to that employed in daily time-series studies of mortality. Gent et al. (23003) used the average of all available ambient monitoring data in a large (6691-square-mile) study area to assess exposures. In these studies, as well as those with similar designs (Naeher et al., 1999) and exposure assessment approaches, major concerns pertain to their ability to accurately ascertain exposure to ozone relative to other co-pollutants. For example, neither design considers exposure heterogeneity within the study area, although it is well known that ozone concentrations will, for example, be reduced in areas with high traffic density due to quenching by NO emissions. Similarly, these designs also do not consider heterogeneity in concentrations of co-pollutants, for example due to traffic density, within the study area.

Neas et al. (1995) used an intermediate design in which a central ambient monitor assigned the same exposure to all study subjects, but with the advantage that these measurements were only applied to a relatively small study area where co-pollutant concentrations were also well characterized. In the setting for this study, Uniontown, Pennsylvania, it is likely that the central monitoring site adequately characterized exposures of the study population, although some small degree of heterogeneity in ambient concentrations is still likely to have resulted in error in characterization of exposure. Ross et al. (2002) also employed a similar approach to exposure assessment. Although implemented in a more complex exposure environment, a somewhat similar study design was that of Romieu et al. (1998) in which the study area (subjects' residence and air monitoring data) was restricted to a specific region within the Mexico City urban area.

An additional issue regarding panel studies and ozone exposure is their ability to consider time-activity patterns. Assessment of activity patterns of study subjects is important given the distinct diurnal patterns of ambient ozone concentrations. The studies of Korrick et al. (1998), Brauer et al. (1996), and Kinney et al. (1996) basically incorporate time activity patterns and the ozone diurnal concentration pattern into the design by measuring outcomes before and during/after daytime periods of elevated ozone concentrations in subjects who, by design, are outdoors during this interval. For example, Brauer et al. (1996) measured lung function before and after daylight work periods in a group of outdoor agricultural workers, Korrick et al. (1998) measured lung function before and after subjects conducted an outdoor hike, and Kinney et al. (1996) measured markers of inflammation in bronchoalveolar lavage samples collected before and after subjects exercised outdoors. Neas et al. (1995), while using central-site ambient monitoring data to assign exposures, also collected individual level data on time-activity patterns and calculated weights to develop individual exposure estimates based upon the proportion of daytime hours spent outdoors. When comparing the weighted versus un-weighted (assuming all children spent the same amount of time outdoors) associations, ozone for peak expiratory flow and evening cough were larger than un-weighted effect estimates. These findings indicate the importance of incorporating time-activity patterns in study design and suggest that failure to adjust for individual-level differences in time-activity patterns, as in the majority of studies using central or averaged monitors (Gent et al., 2003; Mortimer et al., 2000; Naeher et al., 1999; Romieu et al., 1998; Ross et al., 2002), may result in smaller effect estimates.

Panel studies may provide an opportunity to examine ozone exposure and specific parameters of disease activity. For example, Gent et al. (2003) followed 271 asthmatic children under age 12 over 6 months for daily symptoms in relation to ozone and $PM_{2.5}$ The need for regular antiasthma medication was considered an indicator for more severe asthma. Not taking any medication on a regular basis and not requiring a bronchodilator suggested the presence of mild asthma. The ozone effects observed in relation to symptoms were seen only in the medication group. Although the medication use provides an attractive index of disease activity, Gent et al. (2003) did not control for pollen and relative humidity that may be potential confounding factors—it is unclear if these are correlated with ozone concentrations in this study area. This represents another potential limitation of this particular study, although other studies have shown independent effects of ozone and pollen. Finally, the absence of personal monitoring, as previously discussed, is a significant limitation in terms of attributing the effects to ozone.

Controlled human exposure studies

Controlled human studies, or clinical studies, involve exposure of human volunteers to ozone and/or other pollutants under carefully controlled conditions, usually in an exposure chamber, but occasionally via face mask. Activity patterns during exposure are usually conducted according to a protocol, and measurement of minute ventilation (V_E) during exercise for each volunteer and ozone concentration (C) in the chamber during exposure allows accurate and precise calculation of personal exposure. Some noninvasive measures of response such as lung function and respiratory symptoms can be made repeatedly before, during, and after a single exposure while other, more invasive ones such as bronchoalveolar lavage (BAL) or airway reactivity can usually only be performed on one occasion for each exposure. Volunteers usually also undergo an exposure to filtered air (FA) alone, which is otherwise identical to that of the ozone exposure and which serves as a control condition. Comparison of the responses to ozone and FA exposures allows independent estimates of ozone-induced effects without confounding variables from the study itself, such as exercise or diurnal effects.

Nature of evidence

At the time that the 1996 EPA Criteria Document (1996a) was written, the following characteristics of response to ozone exposure had been well established in controlled human studies:

AQ7

- Sufficient short-term ozone exposure caused the following acute reversible effects: (1) lung function decrements (e.g., FEV); (2) induction of respiratory symptoms (e.g., cough, pain on deep inspiration); (3) airway epithelial injury leading to a cascade of events, including increased epithelial permeability, inflammation, and repair; and (4) increased airway reactivity to nonspecific stimuli such as methacholine.
- These effects had all been observed to occur in a dose-response fashion down to 0.08 ppm ozone in one or more studies as a result of 6.6-h exposures with moderate, nearly continuous exercise. They had generally not been studied at concentrations below 0.08 ppm.
- Younger adults were more responsive (for symptoms and FEV) than older adults. Sparse data suggested that for low-level, short-duration exposures, 8–11-year-old children experienced FEV_1 responses that were similar to those of young adults but experienced symptom responses that were absent or smaller. No controlled exposure data were available for children exposed to higher levels of ozone or for durations longer than 2 h. Response was accentuated by exercise and increased duration of response (up to 6.6 h), all other things being equal.
- Otherwise similar, healthy young adults varied in their individual responsiveness (FEV and symptoms) to ozone, but with reproducible levels of response at the individual level. The observed interindividual variability in response became more pronounced and the distribution of individual responses more skewed at higher and higher exposures (i.e., which produced larger mean effects).
- Multiday exposures resulted in an attenuation of the FEV and symptom responses after 3–5 days of exposure, and this attenuation of response lasted for several days to 2 weeks following 5 days of exposure. Other less well-established data suggested that the acute reversible FEV responses of volunteers with asthma may be marginally larger than those without asthma.

Many clinical studies published since the 1996 EPA Criteria Document (EPA, 1996a) have generally been consistent with the information that was well established at that time, and no previously well accepted finding has been refuted. Further studies on asthmatics have generally, but not uniformly, supported the earlier impression that the acute lung function response of asthmatics may be slightly greater than those without asthma. Several new observations that are described below have been made since 1996.

1. Small group mean FEV responses have recently been observed following exposure to 0.06 ppm for 6.6 h in volunteers undergoing moderate, nearly continuous exercise. Over the past 10 years, Adams (2002, 2003, 2006a, 2006b) has published a series of studies of exercising young healthy adults exposed to various concentrations of ozone for 6.6 h that are comparable to the series of studies available for the EPA Criteria Document (EPA, 1996a). His findings for exposures to 0.08 and 0.12 ppm are generally similar to those of these earlier studies both with regard to mean and interindividual variability of FEV response as well as with regard to respiratory symptoms.

   Adams (2006a), for the first time, however, also conducted exposures to 0.04 and 0.06 ppm ozone. There was no meaningful evidence of an effect during the 0.04-ppm exposures, but the data suggest that exposure to 0.06 ppm ozone does result in a small group mean FEV decrement relative to FA exposure, but with ambiguous statistical significance. In this study (Adams, 2006a), 30 healthy young individuals were exposed for 6.6 h to six conditions, including filtered air, 0.06 ppm, and 0.08 ppm, among others. Adams (2006a), seeking differences in patterns of response among the different exposures, utilized a Scheffe post hoc test for controlling study-wide level of alpha while making multiple comparisons among the many data points. This test (which is not particularly powerful for detecting specific differences in the context of large numbers of comparisons) did not identify the response of the 0.06-ppm exposure as statistically different from that of the FA exposure. However, alternative statistical tests suggest that the observed small group mean response in FEV_1 induced by exposure to 0.06 ppm compared to FA is not the result of chance alone. The mean difference in the FEV_1 decrements between the two exposures at 6.6 h was approximately 2.9%, which was statistically different ($p < 0.001$) from 0 when tested using a t statistic without correction for multiple comparisons.

   Further examination of the postexposure FEV data [illegible] and mean data at other time points and concentrations also suggest a pattern of response at 0.06 ppm that is consistent with a dose-response rather than random variability. For example, the response at 5.6 h was similar to that of the postexposure 6.6-h response and appeared to also differ from the FA response. The volunteers in this study did not appear to be more responsive to ozone than volunteers in previous studies as the observed response at 0.08 ppm in this study was similar to that of previous studies. Although of much smaller magnitude, the temporal pattern of the 0.06-ppm response was generally consistent with the temporal patterns of response to higher concentrations of ozone in this and other studies. Responses below 0.08 ppm ozone have not previously been observed, but this finding is not totally unexpected because the previously observed FEV responses to 0.08 ppm were in the range of 6%–9%, suggesting that exposure to lower concentrations of ozone would result in smaller, but real FEV decrements. The EPA reanalysis and reinterpretation of the studies of Adams has been questioned by Adams (2007) and by Smith (2007) in presentations to the Clean Air Scientific Advisory Committee. Thus,

the public health significance of responses at 0.06 ppm ozone is still being debated. The Panel recognizes that uncertainty necessarily surrounds a secondary analysis and the integration of results from a single study in one laboratory with 0.06 ppm ozone exposures and results obtained in studies at higher concentrations by other investigators. Resolution of this uncertainty will require that further research be conducted to clarify the issue. The Panel felt strongly that a double-blinded randomized study with the same subjects exposed to multiple ozone concentrations in the 0.04–0.09-ppm range, with appropriate air controls, would aid in reducing the uncertainty in exposure-response relationships below 0.08 ppm ozone.

2. Several new studies have demonstrated that exposure of individuals with atopic asthma to sufficient levels of ozone produces an increase in specific airway responsiveness to inhaled allergens. The observed responses have only been studied for ozone concentrations above the current NAAQS of 0.08 ppm. The lowest concentration at which this response was observed by Kehrl et al. (1999) following a single exposure was 0.16 ppm during 7.6-h exposures with light exercise. Four consecutive days of exposure to 0.12 ppm for 3 h (Holz et al., 2002) also resulted in an increase in responsiveness to inhaled allergen. These findings, in combination with previously observed effects of ozone on nonspecific airway responsiveness and airway inflammation, support the idea that ambient ozone exposure could result in exacerbation of asthma several days following exposure, and provides biological plausibility for the epidemiologic studies in which ambient ozone concentration has been associated with increased asthma symptoms, medication use, emergency room visits, and hospitalizations for asthma.
3. Some new evidence points to an enhanced inflammatory response in people with asthma exposed to ozone compared to people who do not have asthma (Scannell et al., 1996).
4. Holz et al. (1999) demonstrated individual variability in the inflammatory response and reproducibility of the individual differences, as is the case for FEV responses. Several studies, however, have shown no correlation between the magnitude of individual FEV responses and the inflammatory response, suggesting that different mechanisms are responsible for the two effects Holz et al., 1999; Balmes, et al., 1996; Torres et al., 1997).
5. Evidence suggests that although multiple daily exposures result in attenuation of the FEV and symptom effects for subsequent exposures, some of the epithelial damage/inflammation/epithelial permeability responses do not attenuate during a 5-day exposure (Devlin et al., 1997; Jorres et al., 2000).
6. One study (Folinsbee et al., 1994) suggests that nonspecific airway responsiveness to methacholine does not completely attenuate after 4–5 days of prolonged ozone exposure.
7. One study (Frank et al., 2001) suggests that multiple daily exposures to ozone results in persistent functional small airway changes reflected in lower baseline levels of small airway function over time.
8. Several studies (Adams, 2003, 2006a, 2006b) have confirmed an earlier report (Hazucha et al., 1992) that hour-by-hour responses to prolonged exposures with triangular concentration patterns are not adequately described by the responses to constant concentration exposures with the same overall mean concentration. Rather, responses at intermediate time points are influenced by recent peak concentrations. New exposure-response models based upon a differential equation and a logistic (sigmoid-shaped) function have been developed which accurately describe this hourly FEV response to a wide range of exposure conditions for both constant and variable concentration and activity pattern exposures (Smith et al., 1999; McDonnell et al., 2007). The predictive ability of these models in independent data sets has not been assessed.
9. One study (Gong et al., 1998) found relatively small effects of ozone on several indices of cardiovascular (CV) function among many measured. Responses were similar for volunteers with no CV disease and those with essential hypertension. This is a relatively unexplored area, and the implications of the results of this study for explaining potential CV effects of ambient ozone on the general population are unknown.

Special considerations in evaluating controlled human exposure studies

Controlled human studies are powerful tools for assessing acute, reversible health effects of short-term air pollutant exposure. Because of the random assignment of volunteers to treatment group or the randomization of the order of treatment in a crossover study design, these studies are experimental in design and associations observed in controlled human studies can be considered causal in nature, unlike observational studies in which concerns about confounding and other forms of bias are usually present. Because ozone concentration and minute volume can be measured for each individual, accurate and precise estimates of personal exposure are available for estimating the quantitative relationship between exposure and response. The resulting exposure-response models more accurately reflect true causal relationships, improving risk assessment.

Because pollutant concentrations in a chamber can be controlled, the effects of ozone can be studied directly and independently of the effects of other pollutants in the photochemical mixture that occurs in the ambient air. For logistic reasons (e.g., small sample sizes), these studies are limited in their ability to detect causal effects of ozone on relatively rare, but potentially important events (e.g., emergency room visits for asthma exacerbation). Furthermore, for ethical reasons some segments of the population most likely to be sensitive to ozone exposure (e.g., people with severe asthma) or of great public health concern (e.g., children) are

more difficult to study. In these cases, however, controlled human studies can often provide information that complements or enhances the interpretation of findings in observational studies. The best illustration of this is investigating whether ambient ozone exposure causes or contributes to asthma exacerbation, which controlled human studies cannot directly address. Although people with asthma at greatest risk cannot be subjected to experiments, studies of people with mild asthma reveal that ozone exposure increases airway inflammation and epithelial permeability, nonspecific airway reactivity, and airway responsiveness to inhaled antigen. Because airway inflammation is a hallmark of asthma, increased nonspecific airway responsiveness is associated with elevated likelihood of asthma exacerbation, and increased responsiveness to inhaled antigen is a direct mechanism of exacerbation, the results of these clinical studies indicate that it is biologically plausible that ozone exposure can contribute to asthma exacerbation. These findings support observational studies that associate ambient ozone concentration with asthma exacerbation.

Mechanisms of ozone toxicity

Ozone affects the human respiratory system in several ways. Ozone is a highly reactive gas that is deposited throughout the entire respiratory tract from the airways to the alveoli. Its solubility in water is greater than that of oxygen and its oxidant nature renders it able to react with almost any biomolecule along the respiratory tract. The solubility and reactivity likely account for the reported approximately 40% uptake of inspired ozone by the human nasopharynx (in contrast to SO_2, which is >98%). Dosimetry models predict that the tissue dose of inhaled ozone is greatest at the bronchoalveolar junction, which is the pulmonary region experimentally most sensitive to ozone. Recent ozone bolus studies in humans have confirmed that inspired ozone reaches the distal airways and alveoli of sedentary volunteers, and during exercise ozone penetrates deeper and in greater amounts to the distal lung regions. Thus, ozone can affect the entire respiratory tract but maneuvers such as exercise or oral breathing alter regional deposition of the gas.

Toxicity of ozone is primarily attributed to its reactivity and ozonation of unsaturated fatty acids present in lung lining fluids. Controlled exposure studies in animals and humans have provided a strong scientific basis for understanding ozone toxicity. Changes in lung function in response to ozone have been studied in healthy volunteers, people with asthma, and in individuals with chronic obstructive lung disease, as well as in a number of animal models.

In addition to effects on pulmonary mechanics, exposure to ozone at levels near the current NAAQS causes cellular and biochemical changes in the upper and lower respiratory tracts characteristic of an acute inflammatory response. Respiratory tract inflammation and increased cellular permeability are two of the best-studied biological markers of ozone-induced mechanisms of lung injury in animals, including humans. Clinical studies using BAL show ozone-induced increases in polymorphonuclear leukocytes (PMNs), soluble markers of inflammation and repair, and markers of epithelial permeability (Seltzer et al., 1986; Kehrl et al., 1987; Koren et al., 1989; Devlin et al., 1991; Frampton et al., 1997). Soluble mediators of inflammation (e.g., the cytokines interleukin [IL]-6 and IL-8) as well as arachidonic acid metabolites (e.g., prostaglandin E2 [PGE2], PGF2α, thromboxane, and leukotrienes [LTs] such as LTB4) measured in the BAL fluid in humans exposed to ozone also have bronchoconstrictive properties and may be involved in increased airway responsiveness following ozone exposure.

Animal and human studies suggest that genetic factors can play a major role in responsiveness to ozone. Mice exhibit large intrastrain differences in response to ozone, and a genetic locus confers susceptibility to the ozone-induced influx of PMNs into the lung. Recent evidence suggests that single-nucleotide polymorphisms or null alleles may be risk factors for various diseases. Because ozone is a strong oxidant, it is plausible that genetic variations in phase II antioxidant genes, such as glutathione *S*-transferase, could increase ozone responsiveness. For example, children with an absent or nonfunctioning glutathione *S*-transferase Mu-1 (GSTM1) allele living in Mexico City had greater ozone-induced lung function changes than GSTM1-positive children. Tissue biopsies from GSTM1-null individuals were exposed to ozone in vitro and had significantly increased superoxide dismutase (SOD) expression compared to biopsies from individuals with the wild-type allele. Similarly, chamber studies appear to be demonstrating enhanced responsiveness among GSTM1 ozone-exposed individuals.

In summary, the mechanisms of ozone toxicity in producing relatively acute effects are well understood. The dose to the target tissue, injury to the epithelium with resultant inflammation and increased permeability, and genetic polymorphisms that may confer susceptibility are likely interrelated. These animal and clinical laboratory findings provide plausibility for ozone exposure producing the effects reported in the field and epidemiological studies. However, the laboratory responses are (1) typically observed at levels exceeding those found in the environment; (2) are rapidly reversible; (3) indicate a different mechanism of injury for the inflammatory and functional effects based on disparate responses; (4) are not always unique to ozone, other common air pollutants act through similar mechanisms; and (5) offer little direct insight into potential chronic health effects secondary to ozone exposure.

The most relevant experimental evidence on the effects of chronic exposure to ozone comes from a series of interrelated studies conducted under the sponsorship of the National Toxicology Program (NTP) and the Health Effects Institute (HEI). The exposures were conducted at the Pacific Northwest Laboratories (PNL) of the Battelle Memorial Institute and included experimental observations made by PNL scientists conforming to the standard NTP bioassay. In addition to the observations made by PNL scientists, HEI-supported scientists from other institutions made additional observations that focused on noncancer endpoints in respiratory tract tissue ranging from the biochemical level

to studies of pulmonary function. F344/N rats and B6C3F$_1$ mice, of both genders, were exposed to ozone starting at 6 weeks of age for up to 125 weeks (rats) and 130 weeks (mice). The exposure levels were 0, 0.12, 0.5, and 1.0 ppm ozone for 6 h/day, 5 days/week. The highest concentration was viewed as being the highest concentration that could be tolerated with prolonged exposure.

The details of the studies are documented in reports by the NTP (NTP, 1994) and HEI (Boorman et al., 1995; Harkema et al., 1994; Catalano et al., 1995). The HEI reports are rigorously peer-reviewed and are much more detailed than typical open-literature publications. The three HEI reports cited provide an overview and key results of the ozone studies, there are additional eight reports available. They may be found on the HEI Web site. The survival of both rats and mice were generally similar for all groups. Body weights were generally similar across all groups of rats and mice, with an indication of hypoactivity and lower body weights with the highest ozone exposure concentrations. The most pronounced histological changes, metaplasia and hyperplasia, were observed in the nasal tissues of both rats and mice. There was no effect of 0.12 ppm ozone on nasal structure or function. Alveolar epithelial metaplasia and interstitial fibrosis in the lung were observed with ozone exposure. The pulmonary pathology present with exposure to 0.5 ppm and 1.0 ppm ozone was not present at the 0.12-ppm exposure level. There was no increase in neoplasms in either rats or mice associated with ozone exposure. Extensive pulmonary function tests, similar to the research methods used in human clinical studies, were conducted on the rats. The ozone exposure had little or no measurable impact on lung function. The investigators hypothesized that with prolonged exposure the animals became tolerant to the injurious effects of ozone.

The findings in the mice and rats exposed for prolonged periods of time, approximating the life spans of the two species, to high concentrations of ozone complement the extensive observations with short-term exposures that have tended to focus on acute effects. Species differences in the disposition of inhaled ozone are well recognized and must be considered in extrapolating the findings in mice and rats to humans. In addition, despite the general qualitative similarity of toxicity of many chemicals in laboratory animals and humans, the question remains as to the extent the findings in the mice and rats can be quantitatively extrapolated to humans.

Risk assessment

The risk assessment process

Risk assessment is the process by which analysts summarize the available evidence on health effects in terms that are relevant for policy making. In most applications, including for EPA's NAAQS setting process, the risk assessment is *quantitative*, meaning that the policy-relevant outputs of the risk assessment are numerical estimates of specific health impacts. Risk assessment is the first point in the chain of scientific inquiry on pollutant health effects where analysts take the evidence from clinical and epidemiological studies of responses or associations to estimate implications for the population at large under alternative air quality standards. This requires a number of extrapolations, simplifications, and policy judgments that imply large overall uncertainty in the resulting assessment. As the NRC (2002) has advised, "Although the results of benefit analysis may appear to be less certain, EPA should describe the uncertainty as completely and realistically as possible, recognizing that regulatory action might be necessary in the presence of substantial uncertainty."

A very major shift in the nature of the discussion of the scientific evidence occurs at the risk assessment step. Most of the deliberations surrounding interpretation of the health effects literature center on the question of "whether" certain studies indicate a particular type of physiological response to exposure, and whether an observed response has medical significance. In the risk assessment step, however, EPA treats the question of "whether" as resolved in favor of "yes," and shifts to the question of "how much" public health response occurs and how much those impacts could be reduced by changing the NAAQS. Thus, while much debate continues on whether a causal relationship between ozone and premature mortality lies beneath the studies that have detected a statistical association between the two, EPA's risk assessment presumes this to be the case for the sake of making its numerical estimates. EPA notes this presumption in its documents, but the fact is easily lost to those who do not read the source documents in full, but do read the summaries of the numerical estimates of "lives lost" and potential "lives saved" by tightening the NAAQS.

The resulting numerical health effects estimates are supposed to synthesize and be consistent with all of the available evidence, which may be a combination of toxicological, clinical, and epidemiological studies. In practice, however, most of EPA's risk assessment results are merely extrapolations from a single type of study. In the case of ozone, the estimates of premature mortality and hospitalizations—which are given the greatest focus in subsequent policy-making deliberations—are extrapolated from just a few specific epidemiological studies. The uncertainties introduced in the many types of extrapolation that occur in risk assessment will be discussed below, after a brief summary of the mathematical procedure that EPA employs to extrapolate from epidemiological studies to its estimates of health effects.

An epidemiological study reports whether it detected an association between rates or incidences of a particular health effect (e.g., mortality or hospitalization) and levels of ambient pollution. Actually, there is always some kind of association detected, and the question of interest is whether the numerical estimate of the association is positive (i.e., health effects tended to increase as pollution increased), and what the confidence interval (i.e., the range of variability) is around that numerical estimate. The numerical estimate is often reported as a relative change in risk for a particular amount of change in the monitored ambient pollutant

levels, where a relative risk greater than 1.0 implies that the observed association is positive. A "statistically significant" positive association is one where the confidence range for the numerical estimate of relative risk is always greater than 1.0. Often positive associations are reported as evidence of a health effect, yet such estimates are not statistically significant, which means that the probability that the estimated association is actually negative exceeds the specified confidence level (usually 2.5%). The degree of insignificance can be so large that the probability of the association being negative may reach 50% whereas the point estimate of the risk coefficient remains positive. Nevertheless, EPA uses such nonsignificant associations to generate positive point estimates of health effects from ambient pollution, and of public health gains by tightening the ambient standard.

In the risk assessment, EPA takes the numerical estimate of the association from a particular epidemiological study and interprets that numerical estimate as the slope of a "concentration-response" function for the population at large. The concentration-response function provides an estimate of the health effects risk level of an entire population (e.g., the population of a metropolitan area) when exposed to different levels of a specific measure of ambient air pollution. The pollutant metric used in the concentration-response function is the same as the one measured in the original study (e.g., in terms of the daily average ozone concentration measured at a particular set of monitors in the case of daily ozone mortality studies). Because of the particular shape of the concentration-response relationship that EPA employs, EPA assumes that the change in population risk per unit change in ambient concentrations will be the same at much lower concentrations as it was for the particular levels of pollution that were present at the time the study was conducted. Notably, the concentration-response function also assumes that the level of population risk will be the same for any mix of air pollution that may coexist with a particular value of the single pollutant metric used in the concentration-response function.

EPA then uses recently monitored daily average concentrations of ozone for a specific city to estimate the level of health effects on each day in that city by calculating the "response" level projected by the concentration-response function for each day's ozone level. Daily effects estimates are summed over a full ozone season of daily ozone concentrations to obtain its reported ozone season health effects incidence levels. These estimates are compared to estimates at different levels of pollution, including an estimate of "background" levels, and at estimates of ambient concentrations that would hypothetically occur under alternative NAAQS levels. For the NAAQS standard-setting deliberations, EPA prepares risk estimates only for a selected set of cities, ostensibly because EPA does not wish to extrapolate health effects associations estimated in one particular location (i.e., with a particular socioeconomic mix, weather and lifestyle patterns, and pollutant mixtures) to other locations. In practice, however, EPA makes such extrapolations anyway in how it established the concentration-response functions that it uses for the cities that it does include in its risk assessment. (For example, EPA relies on mortality relative risk estimates for each city included in its risk assessment that were developed as national estimates, using data from a multitude of cities.)

Uncertainty exists in the results of the original studies, but the risk assessment process, which extrapolates from the samples in the studies to a broader population, adds substantially to it—even if that broader population is limited to people in the same location or with the same health conditions as in the sample of the study. The extrapolations required in risk assessment introduce three broad categories of additional uncertainty: (1) extrapolations made to estimate exposure-response relationships, (2) extrapolations made to assess changes in exposures, and (3) policy judgments made regarding "background" pollutant concentrations.

Uncertainty in estimating ambient concentration-response relationships

Uncertainties of the first category—those due to extrapolating from epidemiological evidence to develop a quantified "exposure-response relationship"—have received the lion's share of attention in reviews of EPA's risk assessment. This focus is not because this is the largest source of uncertainty. It is probably because the associated issues are most closely related to the health effects literature, the area in which most of the commentors and advisors surrounding a NAAQS review are experts. The uncertainties are, nevertheless, very large. The discussion of the risk assessment calculations above has already mentioned some of the following extrapolation issues: (1) from the populations in the original studies to current populations in the United States; (2) from levels of pollutants in the original study to current and hypothetical lower levels of concentrations (particularly down to levels that were never experienced in the course of the original study); (3) from the pollutant mix present at the time and place of the original study to the pollutant mix that would be present at different times and locations; and (4) from changes in monitored average concentrations across wide geographical areas to changes in the exposures actually experienced by individuals

In addition, of course, is the fact that the analysis assumes that statistical associations (sometimes statistically nonsignificant ones) reflect a causal relationship with the specific pollutant of concern (i.e., ozone in this case).

Extrapolation is not the only source of uncertainty in developing a quantitative representation of an exposure-response relationship based on epidemiological evidence. As an aside, exposure data for large populations is never available so it is necessary to assess the ambient concentration-response relationship. There are also concerns with quantitative biases in the epidemiological estimates of the relative risk. If a relative risk is estimated for a single pollutant, yet two or more pollutants are affecting health outcomes, then the relative risk for the single pollutant formulation may actually reflect the combined impact of other pollutants that are missing from the statistical model. If they are both exerting an effect,

and both pollutants are positively correlated, the relative risk for the single pollutant in question may be overstated. In fact, a positive relative risk could be attributed to a pollutant that is having no effect on health at all if it is correlated to a harmful exposure that is never included in the regression. Similarly, estimates of relative risk have been shown to vary quite substantially when the statistical model controls for weather and other temporal factors in different ways. There is no objective way to determine the correct form for these statistical controls, so that one has no way of knowing which of the many possible relative risk estimates is the best one—and thus, whenever any single estimate is used in the risk assessment, it is almost certainly a biased estimate.

All of the above concerns in how to move from a qualitative interpretation of the epidemiological literature to the quantitative interpretation required in risk assessment contribute to "model uncertainty." This form of uncertainty is distinct and separate from the statistical variability that surrounds the relative risks that are estimated in the epidemiological studies. Whereas there is statistical variability about the magnitude of the relative risk that is estimated using a particular statistical formula with a specific combination of explanatory variables (or model), model uncertainty relates our lack of knowledge of how to choose the best statistical formula, and also the inherent inability of the statistical techniques to inform us about the shape of the true underlying exposure-response relationship. For example, the statistical techniques and data are limited in their ability to distinguish whether the relative risk per unit of additional concentration varies from lower to higher exposures within the observed data. It is, of course, impossible for these methods to determine such functional shape at levels of exposure that are substantially below those in the dataset—but which are not below those being considered in the risk assessment. Even if the question of functional shape could be addressed satisfactorily for the concentration-response relationship, model uncertainty remains due to the question of how the observed changes in ambient concentrations translate into changes in exposures that people actually experience. All of these issues thus imply a large swath of uncertainty that the measure of statistical variability from the original study does not reflect at all.

Sensitivity analyses—both in the epidemiological estimations and by risk analysts considering alternative extrapolations from estimates in the literature—indicate that model uncertainty regarding the choice of statistical formulation and shape of the ambient concentration-response relationship is much larger than the statistical uncertainty surrounding relative risk estimates obtained for any particular statistical formulation using a single set of data. In practice, however, EPA ignores model uncertainty in its primary risk estimates. EPA presents its primary risk estimate with a "confidence range" that appears to be a statement of uncertainty about its health effects incidence estimates; however, this range reflects only the statistical variability reported for the single relative risk estimate that EPA has chosen to rely on.

Uncertainty in estimating changes in ozone exposures

The second category of uncertainty that is inherent in the extrapolations in EPA's risk assessments for setting a NAAQS is much less widely recognized, but is just as important quantitatively. This is the extrapolation that EPA makes to assess how exposures will be changed if EPA tightens the NAAQS. The way that an alternative standard could reduce health risk is by changing exposure levels, at least on some days. In its risk assessment, therefore, EPA has to estimate how and when exposure levels would change as a result of imposing a different NAAQS level or form of standard. This requires important policy judgments described below. EPA's resulting risk estimates are very sensitive to these judgments.

First, because the risk assessment uses only a concentration-response function, it assumes that all individual exposures change in direct proportion to ambient concentrations at a fixed set of area-wide monitors. Thus, to simulate the change in risks under an alternative standard, EPA estimates how attainment of that standard would affect currently measured average ambient concentrations on the existing network of monitors in each metropolitan area in its risk assessment. The extrapolation technique that EPA applies is called the "rollback" method, because it takes a record of previously monitored ambient levels (e.g., all of the monitored hourly ozone values taken in the city being analyzed during the 2004 ozone season) and lowers ("rolls back") each monitored value by a specific proportion.

In the case of ozone, EPA determines the proportion by which each monitored value in the record is reduced, or rolled back, by estimating the percentage by which the highest values within that record would have to be reduced in order to "just attain" the alternative standard. A related percentage reduction is then applied to every other hourly monitored value in the record, depending on the absolute level of ozone that was measured. The percentage reduction applied to these highest values in the distribution of hourly measurements is simply the initial percentage reduction that EPA determined necessary for attainment. As one moves to lower and lower levels on the distribution, that initial percentage is reduced gradually in such a way that it becomes a zero-percentage reduction just as one reaches the tail of the distribution where the measured ozone level would be zero. EPA calls this a "quadratic rollback" formula because the initial rollback percentage declines in a quadratic fashion along the length of the distribution that is being rolled back. The end result is a new distribution of hourly ozone values that lies to the left of the original actually monitored distribution, with greater reductions on the higher ozone days than on the lower ozone days, but always with quite substantial reductions until one reaches near-zero ozone values.

The new distribution exactly meets the alternative standard in question in terms of its peak levels, but use of EPA's rollback formula also makes a strong assumption in its adopted rollback formula that attainment of the necessary peak conditions will have a very substantial effect on ozone exposures in every single hour of the ozone season. Figure 7 provides an example of the distribution of ozone values in

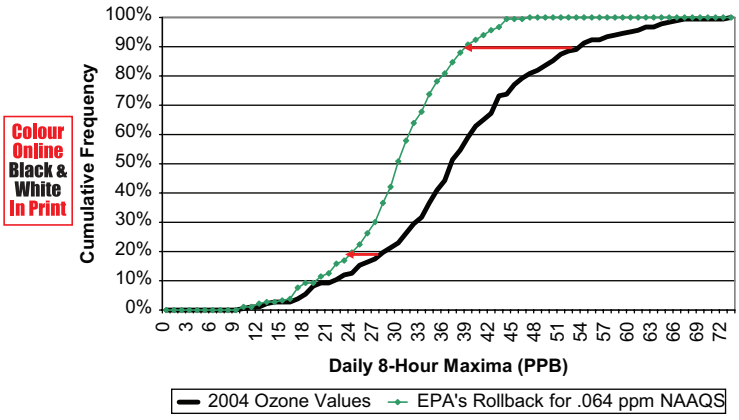

Figure 7. Cumulative distribution function for Detroit 2004 ozone, and distribution after EPA's rollback to simulate "just attaining" a 0.064-ppm ozone NAAQS. (See colour version of this figure online at www.informahealthcare.com/iht)

2004 in Detroit, and of what EPA assumes for its risk assessment would be the distribution of ozone concentrations if Detroit were to just attain an alternative standard of 64 ppb. At each point moving up the graph vertically (which is like moving from the lowest percentile of hourly concentrations to the highest percentile), EPA's estimates of health effects on days at each concentration level are decreased in proportion to the horizontal distance between the original and extrapolated curve.

The total reduction in estimated health effects is (effectively) proportional to the sum of the reductions at each concentration level. For example, consider the 20th percentile concentration level in Figure 7 (i.e., a concentration that is exceeded 80% of the days). The 8-h maximum in 2004 in Figure 7 was about 28 ppb. EPA's rollback method assumes that a NAAQS of 64 ppb would cause all days already at about 28 ppb to be reduced to about 24 ppb. If background levels are assumed to be 20 ppb, this is a 50% reduction, and the portion of risk above background associated with days at that level will also fall by about 50%. Similarly, for days at the 90th percentile ozone levels, the 2004 concentration in Figure 1 was about 53 ppb. EPA assumes concentrations at this level will be reduced to about 38 ppb. If background is 20 ppb, the estimated risk associated with this higher concentration would be reduced by about 45%.

With this knowledge of how the rollback assumptions determine the estimated risk reduction from tightening the standard, one can observe that a large majority of the health effects reductions that EPA has assessed for tightening the ozone standard will be attributable to quite large proportional reductions in ozone in hours when ozone is actually quite low (e.g., almost 90% of the 2004 monitored 8-h average ozone concentrations were less than 50 ppb).

It should be noted here that EPA does not actually compute any benefits for changes in ozone below a value that it calls "Policy Relevant Background," even though it does perform the rollback over all values as shown in Figure 7. The assumption about policy relevant background adds another very important component of uncertainty to EPA's risk estimates and is discussed later. However, regardless of the assumption about Policy Relevant Background, it should be apparent that the assumptions made for using a record of observed ozone concentrations to estimate the pattern of concentrations that would occur under an alternative standard can have an extremely significant impact on the benefits that EPA estimates for such alternative standards. To illustrate this uncertainty further, imagine that an attainment strategy is found that would significantly reduce only the extremely high ozone peaks, such as the highest 5% of the 8-h peak values.

This strategy would also allow the alternative standard to be met. If such a strategy were to be implemented, then the only part of the black line in Figure 7 that would shift as a result of attaining the tighter standard would be the very top segment that lies in the vertical range from 95% to 100%. The rest of the green line in the range from 0% to 95% would lie directly over the black line. Thus, all of the estimates of risk reductions in the lower 95% of the distribution would be eliminated. As a result, estimates of the risk reductions that the alternative standard would provide would be much smaller than EPA has estimated using its assumed quadratic rollback formula. This example is not intended to imply that such an extreme "peak shaving" strategy can be expected in most cases; it is only offered to highlight just how strongly the EPA risk estimates depend on the particular rollback assumptions that it chooses. There is just as much uncertainty in the estimates of risk associated with different possible rollback assumptions (all of which would be consistent with attaining alternative NAAQS standards) as there is uncertainty in the shape and level of the exposure-response curve. This rollback assumption uncertainty is rarely discussed and deserves much more scrutiny in policy deliberations.

Uncertainty associated with EPA's assumption on "Policy Relevant Background"

A third category of uncertainty in the results of EPA's risk assessment for ozone relates to a seemingly innocuous assumption in its risk assessment calculation. EPA calculates and reports risks that are attributable to ozone above the Policy Relevant Background (PRB). The determination of what PRB actually is, however, is extremely difficult. Importantly, uncertainty and debate about the correct PRB value is the single most critical source of uncertainty in the current set of EPA ozone risk estimates. EPA has changed its estimate of PRB since the last ozone review cycle. In 1997, EPA assumed that PRB was 40 ppb throughout the ozone season. In the current review cycle, EPA has estimated PRB for each city using GEOS-CHEM. The new PRB estimates, which vary temporally and by city, are substantially lower than 40 ppb. For example, the maximum hourly value of PRB that EPA now uses for Detroit is 27 ppb, and the seasonal average value over all the hours is only 21 ppb, or approximately half of what it assumed for its 1997 analysis.

EPA changed its PRB assumption with relatively little discussion, but EPA's current estimates of mortality health benefits from tightening the NAAQS depend almost entirely

on this seemingly minor change of assumption about PRB. For example, EPA estimates 24 premature deaths per year in Detroit due to the extent to which its 2004 ("current") ozone concentrations exceed PRB. However, if EPA had used 40 ppb as the PRB assumption, as it did in 1997, it would have estimated less than 0.1 premature deaths per year. (A similar degree of sensitivity applies to risk estimates in the other cities in EPA's risk analysis.) The change in the estimate has nothing to do with uncertainty in the ambient concentration-response relationship, but only because EPA changed this single policy judgment in its risk assessment formula.

It is beyond the scope of this review to determine what the most appropriate assumption for PRB would be; however, the technical and policy debate on what ought to be considered a "policy-relevant" minimum—or the concentrations below which NAAQS policy could not be expected to have any impact—continues to include levels of 40 ppb and even higher. The fact that key health risks associated with ozone (i.e., mortality and hospital admissions) may fall to effectively zero for PRB assumptions within this range of uncertainty reflects a critical source of uncertainty that should be highlighted and not hidden in the communication of risk analysis results. What this really implies is that the sensitivity to the PRB assumption merits further explanation. It more fundamentally reveals that almost all of the health effects that are reported in EPA's risk assessment are attributable to assumed rollbacks in the lower and middle portions of the seasonal distribution of ozone levels. When considering the heath effects that EPA claims could be avoided by applying tighter standards to peak levels of ozone, some individuals logically conclude that those benefits would be due to reductions in exposures to the peak levels of ozone. The sensitivity analysis shows, however, that extremely few of the estimated ozone health impacts are associated with days that have ozone concentrations above the current standard. Thus, for the benefits that EPA's Risk Assessment reports to occur, one must accept two strong but unsubstantiated assumptions that are inherent in EPA's risk assessment methodology: (1) one must accept that EPA's rollback assumption that ozone concentrations will be reduced substantially even on moderate and low ozone days as a result of a standard whose form addresses only the peak days; and (2) one must accept EPA's assumption that the relative risks estimated for higher levels of ozone apply *in equal degree* to changes in exposure to very low levels of ozone.

The foregoing discussion indicates that EPA should communicate more than just its quantitative estimates of health effects. It should also clearly identify the concentrations of ozone that account for the bulk of those estimates and the assumptions about the types of changes in those concentrations that EPA is assuming when it calculates its estimates of the health effects benefits of tightening the current NAAQS.

Need for integrated uncertainty analysis

The difficulties with the way assumptions about PRB affect EPA's risk estimates are just one aspect of a high degree of uncertainty about the magnitude of EPA's risk estimates. In the case of the PRB assumption, almost all of the possible alternative assumptions would lead to very much lower risk estimates. This implies that the uncertainty is not symmetric in both directions around EPA's estimate, but with a systematic bias toward overestimation in EPA's estimates. There are many other important areas of uncertainty in EPA's risk estimates that are also not properly presented or integrated into the analysis, and this fact is another serious concern with EPA's Risk Assessment method.

EPA separates its risk calculations into "primary" and "secondary" risk estimates. Only the primary risk estimates are carried forward into the summary tables of the Staff Paper and other materials likely to be read by the policy community at large, or presented to decision makers such as the Administrator. The primary estimates are calculated using a relative risk coefficient estimated based on a single regression formulation in a single epidemiological study. EPA sometimes reports "confidence bounds" or "uncertainty intervals" for these primary estimates, but *these ranges are based solely on the standard error of that single relative risk coefficient estimate.* In other words, EPA presents a measure of variability as if it were a measure of uncertainty. This may be misleading, especially to a nontechnical audience that may not take the time to learn all of the details of EPA's analysis methods. Thus, it is quite possible that some individuals will believe that these ranges are a reasonably complete representation of the uncertainties about the level of risk. Given that some of these ranges are very wide (in many cases falling into the negative numbers), it would be quite understandable if some individuals were to think these ranges present a comprehensive view of uncertainty, but they do not even start to do so.

As an earlier part of this article has explained, variability is only one very small part of the issue. Uncertainty is what we do not know at all, and which the available data cannot inform. The key uncertainties that EPA does not include in any of the risk estimate ranges that it provides include (1) *Model selection bias* caused by (a) the fact that authors report only one or two of their statistical estimates, and evidence indicates that what is left on the cutting room floor is often of smaller magnitude or less statistical significance; (b) publication bias, wherein studies that find health effects associations are more likely to be published than those that fail to find an effect; (c) EPA's selection of a single paper on which to base its risk estimate; (d) EPA's selection of a single regression from that paper even if the paper reports several. In the case of (d), EPA's pattern has been to select a risk estimate that is the largest and/or most significant from a paper, regardless of the quality of the controls that have been applied to obtain that particular result. For example, EPA uses only risk coefficients from one-pollutant formulations even if a paper contains a two-pollutant formulation. This practice creates a "missing variable bias" that leads to overstatement of the risk; (2) *Model uncertainty* caused by issues such as (a) uncertainty on how to properly control for the confounding effects of time and weather patterns when making statistical estimates of air pollution associations with health; (b)

what the shape of the exposure-response function is given that the risk coefficients are estimated using blunt and noisy ambient concentration-response data; and (3) *Causality*. Even if a positive risk estimate results from many studies, there is still the question of whether any of those associations are evidence of a truly causal effect with ozone such that if ozone were reduced then the health effects that are correlated with ozone will also be reduced. The entire risk assessment starts with a presumption that there is a causal effect, but that does not mean that this is no longer a part of the overall uncertainty in the estimated effects. This is a particularly serious problem for risk estimates that are based entirely on epidemiological evidence rather than clinical evidence, which includes all of the mortality and hospitalization estimates for ozone.

Some, but not all, of the above sources of uncertainty are explored in EPA's "secondary" risk analyses, which are essentially a large volume of impenetrable sensitivity analyses placed in technical support documents and their appendices. The important thing, however, is not to "do" such sensitivity analyses, but rather to use sensitivity analyses to identify key sources of uncertainty and then to *integrate these uncertainties into a single probability distribution over the final, summary risk estimate.* The NRC review of EPA's risk analysis methodology[23] strongly advised that EPA provide primary risk estimates that are founded on such an integrated uncertainty analysis. The ozone risk assessment reflects no such change in approach. This stands as a significant flaw in the credibility of the resulting estimates and their appropriateness as information to support policy decisions.

Just the listing of the key sources of uncertainty provided above suggests that their preponderant effect is to create an upward bias in EPA's risk estimates. Even when leaving aside the uncertainty about causality, analyses produced by other researchers have demonstrated that when an integrated uncertainty analysis is performed, the large majority of probability in the estimates falls far below the primary estimates that EPA reports.

Summary

This paper summarizes critical considerations in evaluating scientific evidence on the health effects of ambient ozone that informed the EPA Administrator's judgment in revising the National Ambient Air Quality Standard from 0.08 ppm to 0.075 ppm with an 8-h averaging time. It is our opinion that these issues will also dominate the next review of the ozone health standard that has just been initiated.

Ozone in ambient air is produced by complex chemical processes from precursors, hydrocarbons, and nitrogen oxides emitted from both natural and man-made sources. Ambient ozone concentrations are also influenced by lightning and periodic intrusions of stratospheric ozone into the troposphere. Ambient ozone concentrations vary by time of day, season, and location across the country. The specific concentrations found at a given location are due to multiple factors, including the concentrations of precursors and other chemicals in the air originating from both biogenic and anthropogenic origins, intensity of solar radiation, temperature and meteorological conditions, both near the monitoring site and at upwind locations that influence the formation, degradation, and transport of ozone. Since promulgation of the original NAAQS in 1971, initially with a 1-h averaging time and since 1997 with an 8-h averaging time, most cities in the United States have made remarkable progress in substantially reducing ambient concentrations of ozone and other criteria pollutants by controlling man-made sources. On average across the United States, the 8-h maximum ozone concentrations have decreased by 21% and the 1-h maximum ozone concentrations have decreased by 29%. As ambient concentrations of ozone decline, the fraction of remaining ozone associated with precursors that are man-made U.S. emissions, the only part that U.S. policy can control, is reduced. This makes further reductions in ambient ozone a challenge in many areas of the United States.

To describe background levels of ozone for making policy judgments on the setting of the NAAQS for ozone, EPA has defined the "Policy Relevant Background" (PRB) as the ozone concentrations that would be observed in the United States in the absence of anthropogenic emissions of precursors (e.g., VOC, NO_2, and CO) in the United States, Canada, and Mexico. The GEOS-CHEM model used to estimate PRB has both low temporal and spatial resolution. Spatial resolution is based on a 2-degree by 2.5-degree grid, which is equal to 138 miles by 173 miles or 24,000 square miles, an area about half the size of the Commonwealth of Pennsylvania. This low resolution raises serious questions as to the relevance of modeled concentrations to actual concentrations of ambient ozone for specific regulated communities. Assumptions made about the Policy Relevant Background have important implications for interpreting the risk assessment for ozone and the extent to which the standard can be attained by control of U.S. man-made emissions. The Panel is of the opinion that there are alternative approaches to describing background ozone concentrations when making policy judgments on the NAAQS for ozone.

Data on the potential health effects of exposure to ambient levels of ozone that should inform policy judgments on the NAAQS are from five types of studies: human clinical studies, three kinds of epidemiological studies, and toxicological studies. The interpretations of these studies are a matter of debate.

The human clinical studies conducted with controlled ozone exposures of exercising human volunteers provide useful information on changes in respiratory function. There is clear evidence of temporary functional changes after protracted exposure to ozone at concentrations of 0.08 ppm and higher. EPA's reanalysis of data from a single clinical study of 30 volunteers suggests that prolonged exposure to 0.06 ppm causes small changes in lung function in some exercising individuals. Some scientists disagree with the EPA's reanalysis and its interpretation. Although these findings have not been confirmed or replicated, the responses to 0.06 ppm ozone in that study are consistent with the presence of an

exposure-response curve with responses that do not end abruptly below 0.08 ppm. The uncertainty that necessarily surrounds a secondary analysis and the integration of results from a single study in one laboratory with 0.06 ppm ozone exposures and results obtained in studies at higher concentrations by other investigators requires that further research be conducted to clarify the issue. For example, a double-blinded randomized study with the same subjects exposed to multiple ozone concentrations in the 0.04–0.09-ppm range, with appropriate air controls, would aid in reducing the uncertainty in exposure-response relationships below 0.08 ppm ozone.

Major long-term epidemiological studies have not shown an association between ozone exposure and long-term mortality. These same studies, which compare the life expectancies of groups of people living in areas with different long-term average pollutant concentrations, provided key evidence of an association between particulate matter and long-term mortality for setting the particulate matter standard.

Time-series analyses consider the association between daily fluctuations in ambient ozone concentrations and day-to-day death rates within a particular city or other locale. These analyses have yielded variable results. Statistically significant positive associations between ozone and mortality have been observed in a small minority of the single-city analyses, with no statistically significant associations observed in most cities, even though some of the studies investigated time periods that extended to several decades ago when ozone levels were much higher than they are today. The National Morbidity and Mortality Air Pollution Study used a Bayesian hierarchical analysis, which is supposed to combine information across cities in an optimal way, to analyze data from nearly 100 cities. The combined analysis of all cities did find a small statistically significant positive association between ozone exposure and mortality. However, only six of the cities showed a statistically significant association between ozone exposure and mortality. The heterogeneity of the individual city-specific results across the United States suggests that a single national ozone concentration-mortality coefficient is not appropriate and its use should be questioned. Time-series analyses have also been performed to examine the relationship between ozone exposure and hospital admissions. However, in contrast to the mortality studies, there have been very few new studies since 1997. Those that have been done imply stronger associations with other pollutants than with ozone. In the near term, it will be important to conduct updated time-series studies that extend to more recent times and, thus, include periods of improved air quality. These studies will be most useful when at least one of the ozone metrics evaluated is the current 8-h averaging time used in the ozone standard.

Panel studies typically measure specific health outcomes repeatedly for a defined group of people for short periods of time and assess how these outcomes are associated with repeated measures of an ambient air pollution mixture that contains ozone. Two studies in which personal exposure to ozone was more accurately assessed than in other panel studies observed associations between low concentrations of ambient ozone and temporary decrements in lung function in individuals engaging in strenuous activities (i.e., agricultural field workers and mountain hikers). Although associations between ambient ozone concentration and asthma status outcomes have been observed in some, but not other panel studies, less precise exposure assessment makes it more difficult to ascribe the effects to ozone with certainty. Such ozone exposure–asthma outcome relationships are, however, biologically plausible, as demonstrated at higher concentrations in controlled studies. The inclusion of an 8-h averaging time ozone metric in future studies will aid in interpreting the results in setting future ozone standards that use this metric.

An enlarging body of toxicological data from human and laboratory animal studies provides a basis for hypothesizing how ozone may cause biological changes with relatively high, short-term exposures in excess of the current NAAQS. Quantitative models do not exist for extrapolation of these short-term observations to lower ozone concentrations typical of ambient levels currently observed across the U.S. Lifetime studies with rats and mice exposed 6 h/day, 5 days/ week to 0, 0.12, 0.5, or 1.0 ppm ozone revealed no difference in life span and only modest effects on body weight associated with ozone exposure. The ozone exposure had little or no measurable effect on pulmonary function. Histopathological changes were observed in the nasal and lung tissues of both rats and mice at the two highest ozone exposure levels. There were no ozone-related increases in the incidence of neoplasms in either species. Species differences in the disposition of inhaled ozone and in responses to inhaled toxicants require caution in quantitative extrapolation of these findings to humans.

During the Workshop discussions, major questions were raised concerning the conduct and reporting of the risk assessment used by the EPA to inform policy judgments in proposing the ozone NAAQS. A principal concern was that the risk assessment failed to integrate the key uncertainties in estimating the health risks of current ambient ozone and levels estimated to occur with alternative standards. The National Research Council previously recommended to the EPA that it integrate key uncertainties in assessing health risks of air pollutants into a single probability distribution when reporting final summary health risk estimates. The calculated health risks, and estimated reductions in risk for alternative standards, are highly dependent on the numerous assumptions used. When risk estimates are known to be highly sensitive to selection of a particular parameter, such as the Policy Relevant Background, then alternative assumptions should be included in the analysis and a range of uncertainty presented. Consideration of alternative approaches to defining Policy Relevant Background will, of necessity, extend to revision and re-interpretation of the risk assessment for ozone.

Another issue in the most recent risk assessment was its focus on ozone concentration-response functions using 24-h

ozone concentrations. Only a very few of the exposure-excess risk coefficients used in the risk assessment were based on 8-h ozone concentrations, the averaging time of the ozone NAAQS since 1997. Thus, the calculated excess risk ascribed to the ozone concentrations measured in 2002, 2003, and 2004 were of uncertain relevance in setting a NAAQS with an 8-h averaging time. This is the case because the relationship between ambient ozone concentration-response coefficients for the three ozone metrics (1-h maximum, 8-h maximum, and daily average) is variable over time for any given city and among cities.

Future risk assessment can also be improved with regard to the manner in which the health risk findings are communicated. The recent risk assessment did not communicate how much of the estimated excess health risks was due to exposures that were on days where the peak 8-h average was below the standard of 84 ppb set in 1997. Moreover, it did not communicate that only a small fraction of the risk reductions that EPA estimates would ensue from a tighter NAAQS is due to changes in the peak ozone concentrations that are the public health concern. Analyses reproducing EPA's risk estimates revealed that a substantial share of the estimated change in health effects when tightening the standard from 84 ppb to a NAAQS in the range of 64–74 ppb was due to EPA's assumption that ozone would change on low-concentration days (e.g., those with peaks of 60 ppb and lower) to yet lower levels that would be unlikely under alternative Policy Relevant Background assumptions.

It is widely known that the science on ozone in ambient air and its health effects is extraordinarily complex. In this paper, we have touched upon key scientific issues that are crucial to the policy judgments that must be made in setting the NAAQS for ozone. In the preceding sections, we have examined much of the science behind these issues in the hope that they will broaden and expand the public discussion around the science undergirding the NAAQS for ozone. We hope that our review will stimulate further discussion of these scientific issues, conduct of additional research, and conduct of new analyses that will provide an improved scientific basis for the policy judgments that will have to be made by future EPA Administrators in considering revision of the ozone standard.

Acknowledgments

The Working Conference held in Rochester, New York, was organized by Professor Mark J. Utell at the University of Rochester. The organization and conduct of the conference were greatly facilitated by capable support from Michael Terry, Patricia Noonan, and Eric Rushton, University of Rochester. Dr. Michael Brauer, University of British Columbia, participated in several sessions of the Conference and his valuable contributions to the discussion and the manuscript are acknowledged. Tom Rickey, University of Rochester, and Dr. Ricki Lewis, CareNet Medical Group, provided valuable editorial assistance. Mildred Morgan, who works with Dr. Roger McClellan, provided valuable assistance in preparing the final report.

***Declaration of interest*:** The Working Conference and preparation of this report were funded by the American Petroleum Institute in the form of a grant to Professor Utell. Each of the participants was reimbursed for their travel expenses and received an honorarium for participation in the Conference. The employment affiliate on each of the participants is shown at the beginning of the report. However, it should be emphasized that each individual's participation was on a professional basis and, thus, the input to the Conference discussions and to this report are solely the responsibility of the individual participants.

References

Adams WC. 2002. Comparison of chamber and face-mask 6.6-hour exposures to ozone on pulmonary function and symptoms responses. Inhal Toxicol 14:745–764.

Adams WC. 2003. Comparison of chamber and face mask 6.6-hour exposures to 0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses. Inhal Toxicol 1, 265–281.

Adams WC. 2006a. Comparison of chamber 6.6-hr exposures to 0.04-0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses. Inhal Toxicol 18, 127–136.

Adams WC. 2006b. Human pulmonary responses with 30-minute time intervals of exercise and rest when exposed for 8 hours to 0.12 ppm ozone via square-wave and acute triangular profiles. Inhal Toxicol 18:413–422.

Adams., WC. 2007. Public Comment to CASAC Ozone Review Panel TeleconferenceMarch52007[www.epa.gov/sab/pdf/pub_comments_03-05-07_dr_wm_Adams_uc-davis.pdf].

Balmes Chen LL, Scannell C., Tager I, Christian D, Hearne PQ, Kelly T, Aris RM 1996. Ozone-induced decrements in fev1 and fvc do not correlate with measures of inflammation. Am J Respir Crit Care Med 153:904–909.

Bell ML, McDermott A, Zeger SL, Samet JM, Dominici F. 2004. Ozone and short-term mortality in 95 US urban communities 1987-2000. J Am Med Assoc 292:2372–2378.

Bell ML, Dominici F, Sarnet JM. 2005. A meta-analysis of time-series studies of ozone and mortality with comparison to the national morbidity mortality and air pollution study. Epidemiology 16:436–445.

Bell ML., Peng RD, Dominici F. 2006. The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations. Environ Health Perspect, ~~in press. doi:10.1289/ehp.8816.~~ Available ~~at http://dx.doi.org/ [accessed 23 January 2006].~~ AQ10

Bell ML, Kim JY, Dominici F. 2007. Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time-series studies. Environ Health Perspect 115:1591–1595.

Boorman GA, Catalano PJ, Jacobson BJ, Kaden DA, Mellick PW, Nauss KM, Ryan LM. 1995. Consequences of prolonged inhalation of ozone on F344 rats: collaborative studies. Part VI. Background and study design. Health Effects Institute, Cambridge MA, Report No. 065-VI.

Brauer M, Brook JR. 1995. Personal and fixed-site ozone measurements with a passive sampler. J Air Waste Manage Assoc 45:529–537.

Brauer M, Blair J, Vedal S. 1996. Effect of ambient ozone exposure on lung function in farm workers. Am J Respir Crit Care Med 154:981–987.

Brauer M, Brumm J, Vedal S, Petkau AJ. 2002. Exposure misclassification and threshold concentrations in time-series analyses of air pollution health effects. Risk Anal 22:1183–1193.

Burnett RT, Brook JR, Yung WT, Dales RE, Krewski D. 1997. Association between ozone and hospitalization for respiratory diseases in 16 Canadian cities. Environ Res 72:24–31.

Burnett RT, Dewanji A, Dominici F, Goldberg MS, Cohen A, Krewski D. 2003. On the relationship between time-series studies dynamic population studies and estimating loss of life due to short-term exposure to environmental risks. Environ Health Perspect 111:1170–1174.

Catalano PJ, Chang L-YL, Harkema JR, Kaden DA, Lat JA, Mellick PW, Parks WC, Pinkerton KE, Radhakrishnamurthy B, Ryan LM, Szarek JL. 1995. Consequences of prolonged inhalation of ozone on F344 rats: Collaborative studies part xi: Integrative summary. Health Effects Institute, Cambridge, MA, Report No. 065-XI. AQ11

Delfino RJ, Coate BD, Zeiger RS, Seltzer JM, Street DH., Koutrakis P. 1996. Daily asthma severity in relation to personal ozone exposure and outdoor fungal pores. Am J Respir Crit Care Med 154:633–641.

Delfino RJ, Murphy-Moulton AM, Burnett RT, Brook JR, Becklake MR. 1997. Effects of air pollution on emergency room visits for respiratory illnesses in Montreal Quebec. Am J Respir Crit Care Med. 155:568–576.

Devlin RB, McDonnell WF, Mann R, Becker S, House DE, Schreinemachers D, Koren HS. 1991. Exposure of humans to ambient levels of ozone for 6.6 hours causes cellular and biochemical changes in the lung. Am J Respir Cell Mol Biol 4:72–81.

Devlin RB, Folinsbee JL, Biscardi F, Hatch G, Becker S, Madden MC, Robbins M, Koren HS. 1997. Inflammation and cell damage induced by repeated exposure of humans to ozone. Inhal Toxicol 9:211–235.

Dominici F, Zeger SL, Samet JM. 2000. A measurement error model for time-series studies of air pollution and mortality. Biostatistics 1:157–175.

Dominici F, McDermott A, Zeger SL, Samet JM. 2002. On the use of generalized additive models in time-series studies of air pollution and health. Am J of Epidemiol 156:193–203.

Fiore AM, Jacob DJ, Bey I, Yantosca RM, Field BD, Fusco AC, Wilkinson JD. 2002. Background ozone over the United States in summer: Origin trend and contributions to pollution episodes. J Geophys Res (Atmos) ~~107(D15):xx-xx~~.

Fiore AM, Jacob DJ, Liu H, Yantosca RM, Fairlie TD, Li Q. 2003. variability in surface ozone background over the united states: implications of air quality. ~~J Geophys Res. 108(D24):19-1-19-12.~~

Folinsbee LJ, Horstman DH, Kehrl HR, Garder S, Abdul-Salaam S, Ives PJ. 1994. Respiratory responses to repeated prolonged exposure to 0.12 ppm ozone. Am J Respir Crit Care Med 149:98–105.

Frampton MW, Morrow PE, Torres A, Cox C, Voter KZ, Utell MJ. 1997. Ozone responsiveness in smokers and nonsmokers. Am J Respir Crit Care Med 135:116–121.

Frank R, Liu MC, Spannhake EW, Mlynarek S, Macri K, Weinmann GG. 2001. Repetitive ozone exposure of young adults: Evidence of persistent small airway dysfunction. Am J Respir Crit Care Med 164:1253–1260.

Gent JF, Triche EW, Holford TR, Belanger K, Bracken MB, Beckett WS, Leaderer BP. 2003. Association of low-level ozone and fine particles with respiratory symptoms in children with asthma. JAMA 290:1859–1867.

Geyh AS, Xue JP, Ozkaynak H, Spengler JD. 2000. The Harvard Southern California chronic ozone exposure study: Assessing ozone exposure of grade-school-age children in two Southern California communities. Environ Health Perspect 108:265–270.

Gong H, Wong R, Sarma RJ, Linn WS, Sullivan ED, Shamoo DA, Anderson KR, Prasad SB. 1998. Cardiovascular effects of ozone exposure in human volunteers. Am J Respir Crit Care Med 158:538–546.

Gryparis A, Forsberg B, Katsouyanni K, Analitis A, Touloumi G, Schwartz J, Samoli E, Medina S, Anderson HR., Niciu EM., Wichmann H-E., Kriz B, Kosnik M, Skorkovsky J, Vonk JM Dortbudak Z. 2004. Acute effects of ozone on mortality from the air pollution and health: A European approach project. Am J Respir Crit Care Med 170:1080–1087.

Harkema JR, Mauderly JL. 1994. Consequences of prolonged inhalation of ozone on F344 rats: Collaborative studies. Part V: Effects on pulmonary function. Health Effects Institute, Cambridge, MA, Report No. 065-V.

Hazucha MJ, Folinsbee LJ, Seal E. 1992. Effects of steady-state and variable ozone concentration profiles on pulmonary function. Am Rev Respir Dis 146:1487–1493.

Holz O, Jorres RA, Timm P, Mucke M, Richter K, Koschyk S, Magnussen H. 1999. Ozone-Induced inflammatory changes differ between individuals and are reproducible. Am J Respir Crit Care Med 159:776–784.

Holz O, Mucke M, Paasch K, Bohme S, Timm P, Richter K, Magnussen H, Jorres RA. 2002. Repeated ozone exposures enhance bronchial allergen responses in subjects with rhinitis or asthma. Clin Exp Allergy 32:681–689.

Huang Y, Dominici F, Bell ML. 2005. Bayesian hierarchical distributed lag models for summer ozone exposure and cardio-respiratory mortality. Environmetrics 16:547–562.

Ito K. 2003. Associations of particulate matter components with daily mortality and morbidity in Detroit Michigan. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Special Report. Boston, MA: Health Effects Institute, 143–156. Available at http://www.healtheffects.org/Pubs/TimeSeries.pdf [accessed 12 May 2004].

Ito K, De Leon SF., Lippmann M. 2005. Associations between ozone and daily mortality analysis and meta-analysis. Epidemiology 16:446–457.

Janes H, Dominici F, Zeger S. 2007. Trends in air pollution and mortality: An approach to the assessment of unmeasured confounding. Epidemiology 18:416–423.

Jerrett M, Burnett RT, Ma R, Pope CA III, Krewski D, Newbold KB, Thurston G, Shi YL, Finkelstein N, Calle I, Thun MJ 2005. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 16:727–736.

Jerrett M, Burnett RT, Pope CA 3rd, Ito K, Thurston G, Krewsi D, Shi Y, Calle E. Thun MJ 2009. Long-term ozone exposure and mortality. N Engl J Med 360:1085–1095.

Jorres RA, Holz O, Zachgo W, Timm P, Koschyk S, Muller B, Grimminger F, Seeger W, Kelly FJ, Dunster C, Frisher T, Lubec G, Waschewski M, Niendorf A, Magnussen H. 2000. The effect of repeated ozone exposures on inflammatory markers in bronchoalveolar lavage fluid and mucosal biopsies. Am J Respir Crit Care Med 161:1855–1861.

Kehrl HR, Vincent LM, Kowalsky J, Horstman DH, O'Neil JJ, McCartney WH, Bromberg PA. 1987. Ozone exposure increases respiratory epithelial permeability in humans. Am Rev Respir Dis 135:1124–1128.

Kehrl HR, Peden DB, Ball BA, Folinsbee LF, Horstman DH. 1999. Increased specific airway reactivity of persons with mild allergic asthma after 7.6 hours of exposure to 0.16 ppm ozone. J Allergy Clin Immunol 104:1198–1204.

Kinney PL, Thurston GD, Raizenne M. 1996. The effects of ambient ozone on lung function in children: A reanalysis of six summer camp studies. Environ Health Perspect 104:170–174.

Koren HS, Devlin RB, Graham DE, Mann R, McGee MP, Horstman DH, Kozumbo WJ, Becker S, House DE, McDonnell WF, Bromberg PA. 1989. Ozone-induced inflammation in the airways of human subjects. Am Rev Respir Dis 139:407–415.

Korrick SA, Neas LM, Dockery DW, Gold DR, Allen GA, Hill LB, Kimball KD, Rosner BA, Speizer FE. 1998. Effects of ozone and other pollutants on the pulmonary function of adult hikers. Environ Health Perspect 106:93–99.

Koutrakis P, Suh HH, Sarnat JA, Brown KW, Coull BA, Schwartz J. 2005. Characterization of sensitive subpopulations living in Baltimore and Boston. Health Effects Institute Research, Cambridge, MA, Report 131:1–65.

Krewski D, Burnett RT, Goldberg MS, Hoover K, Siemiatycki J, Jerrett M, Abrahamowicz M, White WH. 2000. Reanalysis of the Harvard Six Cities study and the American Cancer Society study of particulate air pollution and mortality. A Special Report of the Institute's Particle Epidemiology Reanalysis Project. Health Effects Institute, Cambridge, MA.

Kunzli N, Medina S, Kaiser R, Quenel P, Horak F Jr, Studnicka M. 2001. Assessment of deaths attributable to air pollution: should we base risk estimates on time series or on cohort studies? Am J Epidemiol 153:1050–1055.

Lee K, Parkhurst WJ, Xue J, Ozkaynak H, Neuberg D, Spengler JD. 2004 Personal/indoor/outdoor exposures of children in Nashville Tennessee. J Air Waste Manage Assoc 54:352–359.

Levy JL, Chemerynski SM, Sarnat JA. 2005. Ozone exposure and mortality an empiric bayes metaregression analysis. Epidemiology 16:458–488.

Linn WS, Szlachoic Y, Gong H Jr, Kinney PL, Berhane KT. 2000. Air pollution and daily hospital admissions in metropolitan Los Angeles. Environ Health Perspect 108:427–434.

Lipfert FW, Perry HM Jr, Miller JP, Baty JD, Wyzga RE, Carmody SE. 2000. The Washington University–EPRI Veterans' Cohort Mortality Study: Preliminary results. In: Grant LD, ed. *PM*2000: Particulate Matter and Health. Inhal Toxicol 12:41–73.

Lipfert FW, Wyzga RE, Baty JD, Miller JP. 2006a. Constituents and related air quality variables as predictors of survival in a cohort of U.S. military veterans. Inhal Toxicol 13:645–657.

Lipfert FW, Wyzga RE, Baty JD, Miller JP. 2006b. Traffic density as a surrogate measure of environmental exposures in studies of air pollution health effects: Long-term mortality in a cohort of US veterans. Atmos Environ 40:154–169.

Lippmann M, Ito K, Nádas A, Burnett RT. 2000. Association of particulate matter components with daily mortality and morbidity in urban populations. Health Effects Institute, Cambridge, MA, Research Report No. 95.

Liu SL-J, Koutrakis P, Leech J, Broder I. 1995. Assessment of ozone exposures in the greater metropolitan Toronto area. J Air Waste Manage Assoc 45:223–234.

McDonnell WF, Stewart PW, Smith MV. 2007. The temporal dynamics of ozone-induced Fev1 changes in humans: An exposure-response model. Inhal Toxicol 19:483–494.

Moolgavkar SH, Luebeck EG, Hall TA., Anderson EL. 1995. Air pollution and daily mortality in Philadelphia. Epidemiology 6:476–484.

Mortimer KM, Tager IB, Dockery DW, Neas LM, Redline S. 2000. The effect of ozone on inner-city children with asthma: identification of susceptible subgroups. Am J Respir Crit Care Med 162:1838–1845.

Mortimer KM, Neas LM, Dockery DW, Redline S, Tager IB. 2002. The effects on air pollution on inner-city children with asthma. Eur Respir J ~~1:699–703.~~

Naeher LP, Holford TR, Beckett WS, Belanger K, Triche EW, Bracken MB, Leaderer BP. 1999. Healthy women's PEF variations with ambient

summer concentrations of PM_{10} $PN_{2.5}$ SO_4^{2-} H^+ and O_3. Am J Respir Crit Care Med 160:117–125.

National Research Council. 1983. Risk Assessment in the Federal Government: Managing the Process. Washington, DC: National Academy Press. Available at http://books.nap.edu/catalog.php?record_id=366.

National Research Council.1993. *Issues in Risk Assessment Part* 3: A Paradigm for Ecological Risk Assessment. Washington, DC: National Academy Press. Available at http://books.nap.edu/catalog.php?record_id=2078.

National Research Council. 1994. Science and Judgment in Risk Assessment. Washington, DC: National Academy Press. Available at http://books.nap.edu/catalog.php?record_id=2125.

National Research Council. 1998. Research Priorities of Airborne Particulate Matter. I. Immediate Priorities and a Long-Range Research Portfolio. Washington, DC: National Academy Press.

National Research Council. 1999. Research Priorities for Airborne Particulate Matter: II. Evaluating Research Progress and Updating the Portfolio. Washington, DC: National Academy Press.

National Research Council. 2001. Research Priorities for Airborne Particulate Matter: III. Early Research Progress. Washington, DC: National Academy Press.

National Research Council. 2002. Estimating the Public Health Benefits of Proposed Air Pollution Regulations. Washington, DC: National Academy Press. Available at http://books.nap.edu/catalog.php?record_id=10511.

National Research Council. 2003. Research Priorities for Airborne Particulate Matter. IV. Summary Report. Washington, DC: National Academy Press.

National Research Council. 2007a. Models in Environmental Regulatory Decision Making. Washington, DC: National Academy Press. Available at http://books.nap.edu/catalog.php?record_id=11972.

National Research Council. 2007b. Scientific Review of the Proposed Risk Assessment Bulletin from the Office of Management and Budget. Washington, DC: National Academy Press. Available at http://books.nap.edu/catalog.php?record_id=11811.

National Toxicology Program. 1994. Toxicology and Carcinogenesis Studies of Ozone (CAS No. 10028-15-6) and Ozone/NNK (CAS No. 10028-15-6/64091-91-4) in F344/N Rats and B6C3 F_1 Mice. National Toxicology Program Technical Report Series No. 440. October 1994.

Neas LM, Dockery DW, Koutrakis P, Tollerud DJ, Speizer FE. 1995. The association of ambient air pollution with twice daily peak expiratory flow rate measurements in children. Am J Epidemiol 141:111–122.

O'Neill MS, Ramirez-Aguilar M, Meneses-Gonzalez F, Hernandez-Avila M, Geyh AS, Sienra-Monge JJ, Romieu I. 2003. Ozone exposure among Mexico City outdoor workers. J Air Waste Manage Assoc 53:339–346.

~~Oltmans SJ, et al.~~ 2006. Long-term changes in trophospheric ozone. Atmos Environ 40:3156–3173.

Romieu I, Lugo MC, Colome S, Garcia AM, Avila MH, Geyh A, Velasco SR, Rendon EP. 1998. Evaluation of indoor ozone concentration and predictors of indoor-outdoor ratio in Mexico City. J Air Waste Manage Assoc 4:327–335.

Ross MA, Persky VW, Scheff PA, Chung J, Curtis L, Ramakrishnan V, Wadden RA, Hryhorczuk DO. 2002. Effect of ozone and aeroallergens on the respiratory health of asthmatics. Arch Environ Health 57:568–578.

Sarnat JA, Koutrakis P, Suh HH. 2000. Assessing the relationship between personal particulate and gaseous exposures of senior citizens living in Baltimore MD. J Air Waste Manage Assoc 50:1184–1198.

Sarnat JA, Brown KW, Schwartz J, Coull BA, Koutrakis P. 2005. Ambient gas concentrations and personal particulate matter exposures: Implications for studying the health effects of particles. Epidemiology 16:385–395.

Sarnat SE, Coull BA, Schwartz J, Gold DR, Suh HH. 2006. Factors affecting the association between ambient concentrations and personal exposures to particles and gases. Environ Health Perspect 114:649–654.

Scannell C, Chen L, Aris RM, Tager I, Christian D, Ferrando R, Welch B, Kelly T, Balmes JR. 1996. Greater ozone-induced inflammatory responses in subjects with asthma. Am J Respir Crit Care Med 154:24–29.

Schwartz J. 2005. How sensitive is the association between ozone and daily deaths to control for temperature? Am J Respir Crit Care Med 171:627–631.

Schwartz J, Spix C, Touloumi G, Bacharova L, Barumamdzadeh T, Le Tertre A, Piekarski T, Ponde de Leon A, Ponka A, Rossi G, Saez M, Schouten JP. 1996. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. J Epidemiol Community Health 50(Suppl 1):S3–S11.

Seltzer J, Bigby BG, Stulbarg M, Holtzman MJ, Nadel JA., Ueki IF, Leikauf GD, Goetzl EJ, Boushey HA. 1986. O_3-induced change in bronchial reactivity to methacholine and airway inflammation in humans. J Appl Physiol 60:1321–1326.

Smith MV, McDonnell WF 1999. A statistical model for FEV1 response to arbitrary dynamic ozone exposure conditions. J Air Waste Manage Assoc 49:289–298.

Smith RL. 2007. Public Comment to CASAC Ozone Review Panel Teleconference March 5 2007 [www.epa.gov/sab/pdf/Smith_casac_03-05-07_comment.pdf].

Smith RL, Xu B, Switzer P. 2009. Reassessing the relationship between ozone and short-term mortality in U.S. urban communities. Inhal Toxicol, submi

Thurston GD, Ito K, Kinney PL, Lippmann M. 1992. A multi-year study of air pollution and respiratory hospital admissions in three New York State Metropolitan Areas: Results for 1988 and 1989 Summers. J Expo Anal Environ Epidemiol 2:429–450.

Thurston GD, Ito K. 2001. Epidemiological studies of acute ozone exposures and mortality. J Expo Anal Environ Epidemiol 11:286–294.

Torres A, Utell MJ, Morrow PE, Voter KZ, Whitin JC, Cox C, Looney RJ, Speers KM, Tsai Y, Frampton MW. 1997. Airway inflammation in smokers and nonsmokers with varying responsiveness to ozone. Am J Respir Crit Care Med 156:728–736.

U.S. Environmental Protection Agency. 1996a. Air Quality Criteria for Ozone and Related Photochemical Oxidants. Research Triangle Park, NC: Office of Research and Development. Report Nos. EPA/600/AP-93/004aF-cF.3v. Available from NTIS, Springfield VA, PB96-185582, PB96-185590, and PB96-185608, and at http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_cr_cd.html [accessed 9 November 2004].

U.S. Environmental Protection Agency. 1996b. Review of National Ambient Air Quality Standards for Ozone Assessment of Scientific and Technical Information OAQPS Staff Paper Office of Air Quality Planning and Standards. Research Triangle Park, NC. 27711 EPA-452/R-96-007. June 1996.

U.S. Environmental Protection Agency. 2006. Air Quality Criteria for Ozone and Related Photochemical Oxidants (Final). Washington, DC: U.S. Environmental Protection Agency. EPA/600/R-05/004aF-cF.

U.S. Environmental Protection Agency. 2007a. Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information. OAQPS Staff Paper. EPA-452/R-07-007. July 2007.

U.S. Environmental Protection Agency. 2007b. Ozone Population Exposure Analysis for Selected Urban Areas. EPA 452/R-07-004. January 2007.

U.S. Environmental Protection Agency. 2007c. Analysis of Uncertainty in Ozone Population Exposure Modeling. John Langstaff Memo. January 31 2007.

U.S. Environmental Protection Agency. 2007d. Ozone Health Risk Assessment for Selected Urban Areas. EPA-452/R-07-009. July 2007.

U.S. Environmental Protection Agency. 2007e. National Ambient Air Quality Standards for Ozone: Proposed Rule 40 CFR Part 50. Federal Register 72:37818–37919.

U.S. Environmental Protection Agency. 2007f. Air Quality Trends—Progress Continues in 2006 (last updated June 12 2007). www.epa.gov/airtrends/econ-emissions.html.

U.S. Environmental Protection Agency. 2007g. Regulatory Impact Analyses: 2007 Proposed Revisions to the National Ambient Air Quality Standards for Ground-Level Ozone. Released August 2 2007. www.epa.gov/ttn/ecas/ria.html#ria2007.

U.S. Environmental Protection Agency. 2008. National Ambient Air Quality Standards for Ozone: Final Rule 40 CFR Parts 50 and 58. *Federal Register, ~~March~~* 27:2008.

Vedal S, Brauer M, White R, Petkau J. 2003. Air pollution and daily mortality in a city with low levels of pollution. Environ Health Perspect 111:45–51.

Vingarzan R.2004. A review of surface ozone background levels and trends. Atmos Environ 38:3431–3442.