Texas Public Policy
FOUNDATION

November 9, 2017

Via Federal Express
Mr. Scott Pruitt
EPA Administrator
United States Environmental Protection Agency
EPA Headquarters
Mail Code 1101A
William Jefferson Clinton Building (North)
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

RE: PETITION IN THE MATTER OF: NATIONAL AMBIENT AIR QUALITY STANDARDS FOR PARTICULATE MATTER

Dear Administrator Pruitt:

Enclosed please find our Administrative Petition respectfully requesting the Administrator to reconsider and make less stringent its current national ambient air quality standards ("NAAQS") for fine particulate matter ("PM2.5"), 78 Fed. Reg. 3086 (Jan. 15, 2013), because those standards are based upon faulty assumptions.

Thank you in advance of your careful consideration of the enclosed Administrative Petition.

Respectfully submitted,

Theodore Hadzi-Antich
Senior Counsel
Center for the American Future
Texas Public Policy Foundation

Enclosure

INTRODUCTION

Pursuant to the Right to Petition Government Clause of the First Amendment of the United Sates Constitution,[1] the Administrative Procedure Act,[2] the Clean Air Act,[3] and the United States Environmental Protection Agency's ("EPA's") implementing regulations, Petitioners file this Administrative Petition with EPA's Administrator and, for the reasons set forth herein, respectfully request the Administrator to reconsider and make less stringent its current national ambient air quality standards ("NAAQS" or "standards") for fine particulate matter ("PM2.5"), 78 Fed. Reg. 3086 (Jan. 15, 2013), because those standards are based upon faulty assumptions. Such reconsideration should be part of the current five-year review cycle.

INTEREST OF PETITIONERS

Petitioner Delta Construction Company, Inc. ("Delta") is a California corporation engaged in the business of road construction, performing services such as road paving, reconstruction, shoulder widening, and fabric installation. After 73 years in business, Delta has been forced to close its doors and sell its assets mainly because of regulations governing particulate matter.

Petitioner Dalton Trucking, Inc. ("Dalton") Dalton Trucking, Inc., is a California corporation engaged in the business of operating and leasing loaders, dozers, blades, and water

---

1 "Congress shall make no law . . . abridging . . . the right of the people . . . to petition Government for a redress of grievances." U.S. Const. amend. I. The right to petition for redress of grievances is among the most precious of liberties safeguarded by the Bill of Rights. *United Mine Workers of America, Dist. 12 v. Illinois State Bar Association,* 389 U.S. 217, 222 (1967). It shares the "preferred place" accorded in our system of government to the First Amendment freedoms and has a sanctity and sanction not permitting dubious intrusions. *Thomas v. Collins,* 323 U.S. 516, 530 (1945). "Any attempt to restrict those First Amendment liberties must be justified by clear public interest, threatened not doubtful or remotely, but by clear and present danger." *Id.* The Supreme Court has recognized that the right to petition is logically implicit in, and fundamental to, the very idea of a republican form of government. *United States v. Cruikshank,* 92 U.S. (2 Otto) 542, 552 (1875).

2 5 U.S.C. Section 553(e).

3 42 U.S.C. Section 7401, *et seq.* (sometimes referred to here as the "CAA").

trucks and performs specialized services in open top bulk transportation, lowbed, general freight on flatbeds and vans, as well as rail, intermodal, and 3PL services. Dalton is subject to the PM2.5 standards.

Loggers Association of Northern California, Inc. ("LANC") is a nonprofit California trade association representing the interests of its members involved in the logging industry in Northern California. LANC members are subject to the PM2.5 standards

Robinson Enterprises, Inc. ("Robinson") is a California corporation engaged in various businesses, including forest products and fuels. Robinson is a third-generation family-owned California corporation engaged in harvesting and transportation of forest products, petroleum products, and transportation of various commodities. It has suffered unnecessary financial hardship as a result of various burdensome regulatory requirements, including the PM2.5 standards.

Nuckels Oil Co., Inc. dba Merit Oil Company ("Merit Oil Company") is a California corporation and is a petroleum jobber, wholesaler, and distributor. Merit Oil Company stores, transports, and wholesales a variety of petroleum products, including gasoline, diesel fuels, solvents, and kerosene, and operates a number of delivery trucks and is a family business that has operated in California for three generations. Merit oil Company is subject to the PM2.5 standards.

Western States Trucking Association, Inc. ("WSTA") is a nonprofit California trade association representing the interests of over 1,000 members involved in a variety of business throughout California whose members own and operate on-road and nonroad vehicles, engines, and equipment, which are subject to the PM2.5 standards.

EXECUTIVE SUMMARY

On January 15, 2013, EPA published in the Federal Register a final rule reflecting the

results of its review of its PM NAAQS. 78 Fed. Reg. 3086 (Jan. 15, 2013). The Final PM Rule, with an effective date of March 18, 2013, revised the level of the primary annual NAAQS for PM that is less than or equal to 2.5 microns in diameter ("PM2.5") to 12.0 micrograms per cubic meter ("μg/m3 ") and contained provisions for implementing this standard.

In December 2014, EPA announced the initiation of the current periodic review of the air quality criteria for PM and of the PM2.5 and PM10 NAAQS and issued a call for information in the Federal Register. 79 Fed. Reg. 71764 (December 3, 2014).

"All of the PM NAAQS set to date are based on mass concentration and the assumption that all of the PMs in each size fraction are of equal toxicity on a mass basis. This assumption needs careful review in the current PM review cycle." Roger O. McClellan, *Providing Context for Ambient Particulate Matter and Estimates of Attributable Mortality*, RISK ANALYSIS, 2016; 36(9):1755-1765 at 1757. Recent scientific analyses that cast doubt on the evidence of a causal link between PM2.5 and mortality provide ample reason to reconsider the necessity of the current PM2.5 standards. Given this, the EPA Administrator should not only decline to tighten the primary annual or 24-hour NAAQS for PM2.5, but should consider making the standards less stringent.

As set forth in more detail below, the PM NAAQS should be carefully reconsidered, and the Administrator should open the regulatory process to all interested stakeholders during the current five-year review, including the Petitioners.

STATEMENT OF LAW AND FACTS

I. OVERVIEW OF STATUTORY REQUIREMENTS

The CAA requires the establishment and periodic revision of the PM NAAQS. Section 108 of the CAA (42 U.S.C. § 7408) directs the EPA Administrator to identify and list "air pollutants" that, in his judgment, "cause or contribute to air pollution which may reasonably be

anticipated to endanger public health and welfare" and that the "presence [of which] . . . in the ambient air results from numerous or diverse mobile or stationary sources." He is also required to issue air quality criteria for any air pollutants that are so listed. 42 U.S.C. § 7408(a) & (b). These 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 . . . .**" 42 U.S.C. § 7408(b) (emphasis added). Section 109 (42 U.S.C. 7409) requires the Administrator to propose and issue "primary" (health-based) and "secondary" (welfare-based) NAAQS for pollutants for which air quality criteria are issued under section 108. 42 U.S.C. § 7409(a).

Section 109(b)(1) defines NAAQS primary standards as those that "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." 42 U.S.C. § 7409(b)(1). Section 109(b)(2) provides that secondary standards "shall 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." 42 U.S.C. § 7409(b)(2). Such welfare effects as defined in CAA section 302(h) include "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." 42 U.S.C. § 7602(h).

Section 109(d)(1) of the CAA requires that, at five-year intervals, "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 . . . ." 42 U.S.C. § 7409(d)(1).

Sections 109(d)(2)(A) and 109(d)(2)(B) of the Act require 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 . . . ." 42 U.S.C. § 7409(d)(2).

The Clean Air Scientific Advisory Committee ("CASAC") conducts this review. CASAC has four responsibilities: (1) to advise the EPA Administrator of areas in which additional knowledge is required to assess the adequacy and basis of existing, new, or revised NAAQS; (2) to describe the research efforts necessary to provide the required additional information; (3) to advise the EPA Administrator on the relative contribution to air pollution concentrations of natural and anthropogenic activity; and (4) to advise the EPA Administrator of any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of the NAAQS. Section 109(d)(2)(C).

The purpose of the primary standards is to provide an adequate margin of safety in order to take account of the inherent uncertainties due to inconclusive scientific information, and to provide a measure of protection against dangers not yet identified through research. Through the primary standards, EPA seeks to both prevent pollution levels that have been demonstrated to have adverse effects and to prevent lower pollutant levels that may pose unacceptable risks, even if those risks are, by their nature, not capable of being precisely identified as to their nature or degree. The decision on what approach to take is left to the EPA Administrator's policy judgment. The CAA does not require the Administrator to establish a primary NAAQS which eliminates all risk,

but rather to a level that reduces risk to the extent necessary to protect public health with an adequate margin of safety. *See Lead Industries v. EPA*, 647 F.2d at 1156 n.51 (D.C. Cir. 1980); *Mississippi v. EPA*, 723 F. 3d 246, 255, 262-63 (D.C. Cir. 2013).

In establishing secondary standards, the Administrator must set standards that are neither more nor less stringent than necessary to protect public welfare from any known or anticipated adverse effects associated with the presence of PM. This policy judgment should rely on scientific evidence and analyses about the effects of PM on public welfare, as well as unquantifiable judgments about how to manage uncertainty. The CAA does not require secondary standards be set to eliminate all adverse effects on welfare.

The EPA's task in setting both primary and secondary standards is to establish standards that are neither more nor less stringent than necessary, and it may not consider the costs of implementing the standards, attainability, or technological feasibility. *See generally Whitman v. American Trucking Associations*, 531 U.S. 457, 465-472, 475-76 (2001); *American Petroleum Institute v. Costle*, 665 F. 2d 1176, 1185 (D.C. Cir. 1981).

II. GENERAL SCOPE OF THE CURRENT NAAQS REVIEW

In December 2014, EPA announced the initiation of the current periodic review of the air quality criteria for PM and of the PM2.5 and PM10 NAAQS. 79 Fed. Reg. 17164 (December 3, 2014). The multi-step review process lead to the release of the Final Integrated Review Plan for the National Ambient Air Quality Standards for Particulate Matter ("IRP") in December 2016.

With regard to scope, the current review of the PM NAAQS is focused on the primary and secondary NAAQS for PM2.5 (fine particles) and PM10 (coarse particles). The current primary and secondary PM2.5 standards are meant to protect against the health and welfare effects, respectively, that have been associated with short-term (i.e., hours up to one month) or long-term

(i.e., one month to years) exposures to fine particles. The primary and secondary PM10 standards are meant to protect against the effects associated with exposures to coarse particles. Important aspects of the current review include EPA's assessment of the health and welfare effects that have been associated with short- or long-term exposures to PM based on size fractionated PM mass, with a particular focus on the PM2.5 and PM10-2.5 size fractions. In addition, as in the most recent review, EPA will assess the available scientific evidence for health or welfare effects associated with additional size fractions (e.g., ultrafine particles) and with particular PM components or groups of components, sources, or environments (e.g., urban and non-urban environments).

Based on the available scientific information, EPA is considering the extent to which the current PM2.5 and PM10 standards are requisite to protect public health and welfare, within the meaning of section 109(b) of the CAA. To the extent the available information calls into question the protection afforded by one or more of the existing PM standards, EPA has indicated tha it plans to consider potential alternatives that could be supported by the available scientific evidence and, as available, exposure-/risk-based information, in terms of the basic elements of the NAAQS (indicator, averaging time, form, level).

ARGUMENT

I. THE UNCERTAINTY OF THE SCIENCE REGARDING AMBIENT PARTICULATE MATTER CAUSING ADVERSE HEALTH EFFECTS IS GREATER THAN EPA HAS ADMITTED

> In the United States and some other industrial democracies, where people and their governments tend to be risk averse, legislatures, courts, and administrative entities usually create a presumption favoring more safety rather than less. The definitions of risk in law are often vague ("reasonable certainty of no harm" or "adequate margin of safety") and are likely to encourage an unrealistic belief that risks can be minimized or even eliminated altogether."

- Donald Kennedy, Editor-in-Chief, Science 309: 2137 (30 September 2005)

Roger O. McClellan addresses the scientific evidence relating to NAAQS for PM2.5 in his recent works *Role of Science and Judgment in Setting National Ambient Air Quality Standards: How Low Is Low Enough?*, 5 AIR QUALITY, ATMOSPHERE & HEALTH 243 (2012) (questioning the unbiased nature of EPA NAAQS determinations) (hereinafter, "*Role*") (attached as Exhibit A), and *Providing Context for Ambient Particulate Matter and Estimates of Attributable Mortality*, RISK ANALYSIS, 2016; 36(9):1755-1765 (specifically addressing the PM2.5 NAAQS) (hereinafter, "*Providing Context*") (attached as Exhibit B).

In *Role*, McClellan focuses on EPA's method of setting primary (health-based) NAAQS. *Role* at 243. The Clean Air Act in 1963 and its amendment in 1970 required "the listing of air pollutants that 'may reasonably be anticipated to endanger public health and welfare.'" *Id.* at 244. Subsequent amendments required reevaluation of the NAAQS in 1980 and every five years thereafter. *Id.* EPA also appointed an independent scientific committee called CASAC to conduct peer review for the NAAQS in 1977. *Id.*

When creating a primary NAAQS, 42 U.S.C. § 7409 allows the EPA Administrator discretion to "address uncertainties associated with inconclusive scientific and technical information at the time the Standard is set" to establish an "adequate margin of safety." *Id.* at 245. Congress has also noted that sensitive populations, particularly those with respiratory problems who are regularly exposed to ambient air, should be accounted for. *Id.* Given these criteria, McClellan notes a problem with interpreting the Clean Air Act: though NAAQS are intended to mitigate risk, the Act is unclear about how much mitigation satisfies the law. This may lead some groups to operate under the false assumption that risks from pollution in ambient air can be eliminated. *Id.*

McClellan discusses the politicized nature of such revision. For example, at its creation, the NAAQS for lead were "constrained and informed by the scientific information, but ultimately based on the policy judgment of a politically responsible decision-maker, the EPA Administrator." *Id.* at 246.

Earlier NAAQS were completed through informal rulemaking, which did not provide a sufficient basis for judicial review according to the United States Court of Appeals for the District of Columbia Circuit. *Id.* After that court struck down one of EPA's NAAQS, EPA developed a more rigorous method of documenting their decision-making process for NAAQS and making public their reasoning. *Id.* This reform, which was enacted subsequently by Congress in somewhat modified form in the Clean Air Act Amendments of 1977, Pub.L. No. 95–95, § 305, 91 Stat. 685, sacrificed speed in rulemaking but improved transparency, McClellan notes with approval. *Id.* at 247.

In 1997, EPA chose to set a separate PM2.5 standard for the first time. Prior to that time, PM 2.5 had been included under the standards for ambient particulate matter under 10 microns (PM10). *Id.* Discussions surrounding the first PM2.5 NAAQS were "very contentious" as the scientists on the committee had "a range of views" so complex that it took a table to diagram them. *Id.* This disagreement was magnified by the D.C. Circuit's decision in *Am. Trucking Assn. v. U.S. EPA*, 175 F.3d 1027 (D.C. Cir. 1999). That decision vacated the 1997 PM10 standards largely because they included the PM2.5 standards. Further, they determined that while EPA's factors used to determine degrees of public health concern related to pollutants were "reasonable," EPA lacked any clear criterion for determining NAAQS. However, the EPA Administrator was not allowed to consider the cost of implementing NAAQS when setting them. *Role* at 247.

The Supreme Court affirmed the basic holding of *Am. Trucking* two years later in *Whitman v. Am. Trucking*, 531 U.S. 457 (2001). Writing for the majority, Justice Breyer clarified further that "§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.'" *Id.* at 494. This sought to solve the problem posed by the Clean Air Act's risk-avoidance language: the EPA Administrator has flexibility to avoid setting standards that chill industry activity and determine "the acceptability of small risks to health." *Id.* Thus the EPA Administrator does not have to set NAAQS that aim at completely eliminating pollutants, as if such a thing were possible. Breyer's opinion allows the Administrator to make his determinations about what level of protection and risk is "adequate" based on his policy judgments when crafting primary and secondary NAAQS.

McClellan states that a "paradigm shift" took place as the amount of scientific evidence regarding pollution's health effects grew. *Role* at 248. Originally, lacking human studies on the effects of pollution on health, scientists agreed that the lowest level at which pollution could be determined "statistically significant" in laboratory animal studies served as the highest level for the "adequate margin of safety." *Id.* at 248-49. (As an aside, in recent years, the wisdom of taking lab animal studies as determinative on this matter has been called into question, and EPA has introduced a factor in its NAAQS calculations that supposedly accounts for this discrepancy. *Id.* at 249.) This decision assumed that certain non-cancer health issues had a linear exposure-response relationship to certain pollutants, an assumption which McClellan discusses further in his analysis. *Id.* McClellan also notes the folly of EPA's initial inclination to "identify levels where an increase in effects is observed and then set the Standard at a lower level." *Id.* Eventually, EPA began linking their standards to pollutant concentrations averaged over multiple years. *Id.* This shift in the statistical forms underlying NAAQS produces challenges when certain studies fail to

provide metrics for their data that would aid EPA in averaging. This difficulty "results in extremely stringent Standards that at best are only very loosely related to the underlying data." *Id.*

McClellan points out that EPA's assumptions about appropriate background levels for certain pollutants, combined with ongoing acceptance of a possibly flawed statistical model for NAAQS, has hamstrung the agency's ability make NAAQS that reflect reality. *Id.* at 250. EPA has assumed that its practice of categorizing concentrations of pollutants above the NAAQS in a linear manner, rather than determining "whether there is a threshold level below which the coefficient for excess risk does or does not hold." *Id.* EPA's insistence on this point has extended to estimating adverse health attributable to each pollutant "down to background concentrations." *Id.* While admitting that he was originally in favor of this approach, McClellan did not expect that advocates of such quantification would take their measurements as "highly accurate projections . . . sometimes without any indication of uncertainty." *Id.*

Due to these statistical challenges, McClellan concludes that "decisions on the selection of specific levels and averaging times for the NAAQS are policy judgments properly reserved to the Administrator informed by the available scientific knowledge." *Id.* at 249. In other words, the implications of Breyer's opinion in *Whitman* extend to the statistical modeling underlying the NAAQS determination. EPA's unreasonable decision to adopt linear modeling, in contravention of *Whitman*'s directive that the Clean Air Act recognizes the need for policy judgment within its "adequate margin for safety" parameter is the paradigm shift McClellan previously mentioned.

McClellan then discusses the PM2.5 indicator. He participated in initial CASAC discussions on the first PM2.5 NAAQS in 1997. He noted that the committee members in large part wished to create a NAAQS that "would mandate the monitoring of PM2.5," but also expressed reservations about setting the NAAQS too stringently given the "absence of convincing data on

PM2.5." *Id.* at 251. He states that the Administrator's initial annual NAAQS on PM2.5 was too stringent and "very precautionary," while the 24-hour NAAQS was less so. *Id.* CASAC's revision of this standard in 2005 recommended a tightening of both standards, with significant pressure to provide unanimous approval. McClellan believed this tightening "was not a scientific decision, but rather a matter of policy judgment that should be left to the discretion of the Administrator." *Id.* He and another colleague did not join CASAC's recommendation. The Administrator tightened the 24-hour NAAQS while leaving the annual one where it was. *Id.* McClellan makes it clear that it is "not appropriate for CASAC to recommend a bright line upper bound on the NAAQS," because that recommendation involves policy judgment beyond scientific analysis. *Id.* at 252. While the Administrator is authorized to make decisions about what constitutes appropriate risk and incorporate it into his standard-setting, the CASAC's narrow job is to provide the Administrator with scientific information that will factor into his final decision. *Id.*

McClellan next addresses the call for "sound science" to inform the Administrator's standard-setting decisions. He agrees wholeheartedly, and supports in principle the efforts of advocacy groups and NGOs to synthesize and submit helpful data for EPA's NAAQS process. *Id.* at 254. However, McClellan heavily criticizes the inclination of some groups to hold certain data as "true" or "false" based on who funded the study that produced the data, and expresses concern about the implicit expectations that "sound science" can provide perfect NAAQS:

> Sound science does not in and of itself make for sound decisions. . . . [S]cience alone cannot identify an acceptable level of health risk, since such levels inherently represent a policy judgment call. Sound science can only inform what are ultimately policy judgments or political decisions. This is especially the case for the setting of NAAQS, in the absence of a clearly defined threshold, which involve decisions as to acceptable health risks which are linked to the level (and form) of the Standard.

Id.

McClellan concludes that while *Whitman* allows the Administrator to set NAAQS in a way that accounts for policy judgment, CASAC itself may not exercise the same judgment in making its recommendations. Instead, McClellan wants CASAC members to draw on their diverse expertise to interpret and distill the vast quantity of scientific data on pollutants. *Id.* at 255. Most notably, McClellan believes that the Administrator would greatly benefit from CASAC's input on "the multiple factors that influence morbidity and mortality from respiratory and cardiovascular disease, the major health outcomes for key criteria pollutants." *Id.* at 256. He reaffirms that if Administrators seek to use the CASAC's unwarranted offering of acceptable ranges as scientific cover for their own political judgments, such action would "transform the Clean Air Scientific Advisory Committee into a de facto Clean Air Standards Setting Committee," a result not intended by Congress in enacting the Clean Air Act. *Id.*

Moving on to McClellan's 2016 paper, he specifically addresses PM2.5 NAAQS in light of new research, analyzing the extent to which PM2.5 may or may not contribute to increased mortality based on the new findings. *Providing Context* at 1755. McClellan takes time to summarize the methodology of each study. Two of the four considered studies incorporate alternative methods of measuring acceptable levels of PM2.5, rather than or in addition to the commonly accepted linear concentration-response modeling that McClellan criticized in his 2012 paper. *Id.* at 1756.

In the following section, McClellan points out that in 2012, the Administrator revised the tightened the primary annual NAAQS for PM2.5 to 12μg/cubic m. The 24-hour standard held steady. *Id.* at 1757. McClellan notes that both of these standards "are based on mass concentration and the assumption that all of the PMs in each size fraction are of equal toxicity on

a mass basis." *Id.* Based on new evidence, McClellan suggests that "this assumption needs careful review in the current PM review cycle." *Id.*

McClellan begins his examination of the relation between PM2.5 and mortality by referencing a major long-term study on the subject called the Harvard Six Cities Study. It measures "changes in ambient PM2.5 concentrations in . . . six cities from the mid 1970s through 2009." *Id.* The study demonstrates a sizable and steady decline in ambient PM2.5. *Id.* at 1757-58. McClellan next notes that the crude and age-adjusted death rates have seen marked improvement in the same time frame. *Id.* at 1758. He includes another table indicating the causes of death for the United States in 2010. *Id.* at 1759. This table lists heart diseases as the most common cause of death, followed closely by cancer. Chronic lower respiratory diseases are a distant third. *Id.* Overall, "it is widely acknowledged today . . . that the regulatory programs grounded in the CAA have had widespread positive impact" in terms of improved air quality. *Id.* This brings up the obvious question of whether current air quality requires stricter primary NAAQS for PM2.5. Such a question hinges on whether PM2.5 is still a significant cause of adverse health effects, which McClellan next examines.

McClellan explains that EPA has a five-level hierarchy (ranging from "causal relationship" to "not likely to be causal relationship") to classify the weight of evidence regarding the relation between a given pollutant and a health hazard. *Id.* at 1760. Notably, this level-based system does not speak to whether current PM2.5 levels in the United States increase the incidence of adverse health effects "over and above baseline rates." *Id.* Even more seriously, this system does not establish whether any given ambient PM2.5 concentration has "a causal attributable effect on health outcomes," including an increase in mortality rates *simpliciter*. *Id.*

Many scientists incorrectly believe the conclusions of EPA's level-based system bears some sort of implication for ambient PM2.5 concentration measurements. *Id.* McClellan faults the authors of the four new studies his paper examines for making a related assumption. One examined study implies that the correlation between PM2.5 levels and excess risk of adverse health effects is reliable no matter the examined concentration and risk level – a proposal with which McClellan expresses reservations. *Id.* at 1760. He also questions why the studies failed to question the EPA Administrator's reasoning in lowering primary annual PM2.5 NAAQS so drastically in 2012. In that instance, the Administrator considered a limited range of data in available studies as reliable evidence of a causal relationship between long-term PM2.5 exposure and increased general death rates. *Id.* at 1761. This conclusion conflicts with the conclusion of all four researchers, who considered all data in their studies to be reliable. *Id.* Since data at all concentrations did not show an equal causal relationship between long-term PM2.5 exposure and increased all-cause mortality, this is a serious omission. The Administrator also entirely failed to take into account the Six Cities Study, because it had not released numbers for PM2.5 as recently as other studies. *Id.* at 1760. McClellan calls the contrast between the Administrator's judgments and the seeming conclusions of the most reliable recent studies on PM2.5 "a critical issue at the interface between scientific information and policy choices." *Id.* at 1761.

McClellan criticizes the four studies at issue further, noting that even though the data does not necessarily support the conclusion that low concentrations of PM2.5 cause an increase in death rates, none of the studies discuss this fact. *Id.* "[T]he official assumption in the last EPA review that all PM2.5 is of equal toxicity on a mass basis," McClellan notes, is especially important in a modern context, when most PM results not from direct emissions but "secondary reactions and associated changes in the chemical and size composition of PM." *Id.* Very little

data that differentiates between directly emitted and secondarily derived PM exists. Such data is necessary to determine whether a mortality increase still correlates with both kinds of PM, and in what concentrations. *Id.* While one study has a more extensive discussion of causality than others, McClellan calls its assumptions "simplistic and . . . naïve" for oversimplifying the way that outside stressors cause an increase in mortality. *Id.* He especially finds the study's skepticism about a PM2.5 range of exposure where no mortality risk exists "unjustified," especially since the authors' own methods of measurement require them to "control for all other risk factors potentially associated with the disease endpoint of concern." *Id.* at 1762. These risks are manifold and complex.

In fact, McClellan reveals, there is "a growing body of evidence of a lack of influence of ambient PM2.5 concentrations on mortality." *Id.* In some states, like California, the risk of increased mortality associate with PM 2.5 has decreased to the point of non-demonstrability. *Id.* Moreover, "[i]t is well recognized by scientists and clinicians . . . that none of the individual cases carry "markers" or any characteristics that allow PM2.5 attributable cases to be distinguished from cases that are attributable to a myriad of other causes." *Id.* Because deaths are only attributed to PM2.5 "on a statistical and population basis," we have no hard evidence of any mortality increase directly attributable to PM2.5. *Id.* The authors of the studies reviewed by McCllelan do not discuss whether more well-documented risks could contribute to or account for increases in mortality currently attributed to PM2.5. *Id.* Given the complexity of determining what risk factors contribute to any given death (and the variance of contribution depending on time, place, and exposure level), this omission is glaring.

McClellan suggests that "an expanded presentation of results" incorporating the Six Cities Study and exposure-response measurements would be more informative to future decision-

making about PM2.5 NAAQS. *Id.* at 1763. He also suggests including baseline population and mortality data to provide context for such determinations. *Id.* at 1764.

Regarding the most current models and studies on PM2.5, McClellan concludes that their estimates are "more likely to overestimate than underestimate the true PM2.5 attributable mortality." *Id.* He also wonders whether the data on mortality attributable to certain PM2.5 concentrations have been skewed by the exposure of certain individuals born in or before the 1970s to PM2.5. *Id.* While he agrees that it is possible that improvements in air quality contributed to reduced mortality, "the impact of PM2.5 reductions is likely very small and difficult to tease out from the myriad of other factors that were likely involved" in this reduction, like widespread improvement in overall socioeconomic status. *Id.*

McClellan is not the only scientist to question the evidence of a significant link between fine particulate matter and mortality rates. James E. Enstrom's paper, *Fine Particulate Air Pollution and Total Mortality Among Elderly Californians, 1973-2002*, INHALATION TOXICOLOGY 2005; 17:803-816, (attached as Exhibit C), found no relationship between levels of fine particulate matter (PM2.5) and mortality. Enstrom's analysis used proportional hazards regression and, adjusting for age, sex, cigarette smoking, and other potential confounding variables, found that "[t]hese epidemiologic results do not support a current relationship between fine particulate pollution and total mortality in elderly Californians, but they do not rule out a small effect, particularly before 1983." *Id.* at 803. Enstrom's research was based on 118,094 Californians enrolled in the American Cancer Society's first Cancer Prevention Study. "For the initial period, 1973–1982, a small positive risk was found: RR [relative risk of death] was 1.04 (1.01–1.07) for a 10-μg/m3 increase in PM2.5. For the subsequent period, 1983–2002, this risk was no longer present: RR was 1.00 (0.98–1.02). For the entire follow-up period, RR was 1.01 (0.99–1.03)." *Id.*

at 803.

Similarly, Enstrom's recent paper, *Fine Particulate Matter and Total Mortality in Cancer Prevention Study Cohort Reanalysis*, DOSE-RESPONSE: AN INTERNATIONAL JOURNAL January-March 2017:1-12, (attached as Exhibit D), independently analyzed the findings in the 1982 American Cancer Society Cancer Prevention Study (CPS II), which had earlier found a positive relationship between PM2.5 and total mortality (and has been the basis for EPA's PM2.5 NAAQS levels). Enstrom used Cox proportional hazards regression on the original questionnaire data, examining results obtained from 292,277 participants in 85 counties with 1979-1983 EPA Inhalable Particulate Network PM2.5 measurements, as well as for 212,370 participants in the 50 counties used in the original 1995 analysis. The 1982 to 1988 relative risk (RR) of death from all causes and 95% confidence interval adjusted for age, sex, race, education, and smoking status was 1.023 (0.997-1.049) for a 10 mg/m3 increase in PM2.5 in 85 counties and 1.025 (0.990-1.061) in the 50 original counties. The fully adjusted RR was null in the western and eastern portions of the United States, including in areas with somewhat higher PM2.5 levels, particularly 5 Ohio Valley states and California. Enstrom concluded there was no significant relationship between PM2.5 and total mortality in the CPS II cohort was found when the best available PM2.5 data were used. Contrary to the original 1995 analysis's finding of a positive relationship by selective use of CPS II and PM2.5 data Enstrom found that the underlying data raises serious doubts about the CPS II epidemiologic evidence supporting the PM2.5 NAAQS.

There have also been relevant contributions to a recent issue of *RISK ANALYSIS.* Anne Smith's paper illustrates the use of alternative approaches to calculating the expected benefits of reducing the NAAQS for PM2.5 from 15 to 12 μg/m3. Anne E. Smith, *Inconsistencies in Risk Analyses for Ambient Air Pollutant Regulations*, RISK ANALYSIS, 2016; 36(9):1737–1744

(attached as Exhibit E). Smith describes the inconsistency between the health risk analysis that EPA uses to support its NAAQS standards and in the Regulatory Impact Analyses (RIAs) related to each NAAQS rulemaking. Risk estimates are prepared during the process of setting the NAAQS level using statistical relationships between measured pollutant concentrations and effects on human health. The final risk estimates are not directly used to set the NAAQS level, but are incorporated into a rationale for the standard intended to show compliance with the statutory requirement that the primary NAAQS protect the public health with a "margin of safety."

In a separate process, EPA relies on the same risk calculations to prepare estimates of the health benefits of the rule that are reported in its RIA for the standard. Although NAAQS rules and their RIAs are released simultaneously, the rationales used to set the NAAQS have become inconsistent with their RIAs' estimates of benefits, with very large fractions of RIAs' risk-reduction estimates being attributed to populations living in areas that will already be attaining the respective NAAQS.

Smith's paper explains the source of this inconsistency and provides a quantitative example based on the 2012 revision of the PM2.5 primary NAAQS. Smith shows that the total risk reduction estimate (avoided premature deaths in 2020) for two approaches. The first was the traditional approach used by EPA in developing RIAs, which assumes deaths are avoided regardless of the ambient concentrations of PM2.5. The analysis in the RIA showed 456 avoided deaths with one concentration–response function using the American Cancer Society cohort and 1,034 avoided deaths using the concentration–response function from the Six Cities Study. Smith also gave lower estimates based on the rationale that EPA used in the latest revision of the NAAQS for PM2.5, with the number of residual avoidable deaths reduced to 21–48, dependent on the concentration–response function used. "The result is that the RIA benefits are substantially

overstated compared to those that would more appropriately reflect the subjective weights expressed by EPA in its rationale for setting the standard at 12 μg/m3." *Id.* at 1741.

Smith finds that a large majority of EPA's estimated health benefit from the 2012 PM2.5 NAAQS are attributable to reductions of PM2.5 in areas that were already in attainment of the PM2.5 NAAQS. RIA calculations of risk reduction in areas already attaining the new NAAQS are given the same weight (i.e., subjective confidence level) as projected benefits from areas that would be exceeding the NAAQS. These RIA calculations are based on assumptions that are inconsistent with the rationale for that NAAQS. This causes RIAs' benefits estimates to be much more substantial than estimates of the expected benefits that could be reasonably inferred from EPA's NAAQS-setting rationale. The overstatement becomes nearly 100% for co-benefits from criteria pollutants in RIAs for non-NAAQS regulations. *Id.* at 1742-43.

Tony Cox was invited to comment on Smith's paper (as well as other papers). Cox points out the flaws in existing models purporting to predict how future changes in exposure to PM2.5 affect mortality. Louis Anthony Cox, Jr., *Rethinking the Meaning of Concentration-Response Functions and the Estimated Burden of Adverse Health Effects Attributed to Exposure Concentrations,* RISK ANALYSIS, 2016; 36(9):1770-1779 (attached as Exhibit F). Basically, the modeling choices affect the concentration-response relations, but equally good varying choices lead to conflicting conclusions regarding any adverse effect from a given level of PM2.5 on mortality. This means that currently available data has questionable efficacy in predicting how future changes in PM2.5 concentrations will affect human health. *Id.* at 1770-75.

The reduced-form regression models used to attempt to establish associations between particular PM2.5 levels and mortality are flawed, but Cox believes that other methods of modeling risk, from simulation to causal Bayesian networks, could be more efficacious in determining

changes in responses from changes in exposure level. *Id.* at 1775-77. Given the flaws in the current data used by EPA, and the possibility of more accurate models as outlined in Cox's paper, it would be irresponsible for EPA to tighten the PM2.5 NAAQS.

The analyses of McClellan, Enstrom, Smith, and Cox provide more than enough reason to reconsider the necessity of the current extremely stringent PM2.5 standards. Given that the causal link between PM2.5 and mortality is tenuous at best and indemonstrable at worst, the EPA Administrator certainly should not tighten the primary annual or 24-hour NAAQS for PM2.5; rather, the Administrator should consider making the standards less stringent.

II. EPA Has Inherent Authority to Reconsider the PM NAAQS

"Agencies are free to change their existing policies as long as they provide a reasoned explanation for the change. When an agency changes its existing position, it need not always provide a more detailed justification than what would suffice for a new policy created on a blank slate. But the agency must at least display awareness that it is changing position and show that there are good reasons for the new policy." *Encino Motorcars, LLC v. Navarro*, 136 S. Ct. 2117, 2125–26 (2016) (internal citations and quotation marks omitted). Furthermore, "[a]n initial agency interpretation is not instantly carved in stone [although] reasoned decision-making ordinarily demands that an agency acknowledge and explain the reasons for a changed interpretation. But so long as an agency adequately explains the reasons for a reversal of policy, its new interpretation of a statute cannot be rejected simply because it is new." *Verizon v. FCC*, 740 F.3d 623, 636 (D.C. Cir. 2014). Accordingly, EPA is free to reconsider its prior decisions on PM NAAQS.

As the Supreme Court has observed, "[a]gency inconsistency is not a basis for declining to analyze the agency's interpretation under the *Chevron* framework. . . . [I]n *Chevron* itself, this Court deferred to an agency interpretation that was a recent reversal of agency policy."). *Nat'l*

Cable & Telecomm. Ass'n v. Brand X Internet Servs., 545 U.S. 967, 981-82 (2005) (citing *Chevron v. NRDC*, 467 U.S. 837, 857-58 (1984).

Accordingly, EPA may determine in connection with the current five-year review as a matter of policy that the PM NAAQS should be made less stringent in light of new scientific studies relating to harm to human health from PM and the new Administrator's policy judgment in evaluating the uncertainties of the evidence. *See Smiley v. Citibank (South Dakota), N. A.*, 517 U.S. 735, 742 (1996) ("[regulatory] change is not invalidating. . . ."); *Van Hollen, Jr. v. Fed. Election Comm'n*, 811 F.3d 486, 496 (D.C. Cir. 2016) ("An agency 'must consider varying interpretations and the wisdom of its policy on a continuing basis.'") (quoting *Brand X*, 545 U.S. at 981). Therefore, EPA is free to revisit the PM NAAQS based upon the instant Administrative Petition.

CONCLUSION

For these reasons, Petitioners respectfully request that, during the current five-year review, the Administrator reconsider the NAAQS PM2.5 standards in light of the issues brought to his attention in this Administrative Petition. The Petitioners also request that they be provided with the opportunity to actively participate in the five-year review as stakeholders with a keen interest in the outcome.

DATED: November 9, 2017

Respectfully submitted,

Theodore Hadzi-Antich
tha@texaspolicy.com
Ryan D. Walters
rwalters@texaspolicy.com

TEXAS PUBLIC POLICY FOUNDATION
901 Congress Avenue
Austin, Texas 78701
Telephone: (512) 472-2700
Facsimile: (512) 472-2728

ATTORNEYS FOR PETITIONERS

cc: Neomi Rao (via Federal Express)
Administrator
Office of Information and Regulatory Affairs
Office of Management and Budget
725 17th Street, N.W.
Washington, DC 20503

Ted Boling (via Federal Express)
Acting Director
President's Council on Environmental Quality
722 Jackson Place, N.W.
Washington, DC 20506

Sarah Dunham (via Federal Express)
Acting Assistant Administrator
Office of Air and Radiation
Mail Code 6101A
USEPA Headquarters
William Jefferson Clinton Building
1200 Pennsylvania Ave., N.W.
Washington DC 20460

EXHIBIT A

Air Qual Atmos Health (2012) 5:243–258
DOI 10.1007/s11869-011-0147-2

Role of science and judgment in setting national ambient air quality standards: how low is low enough?

Roger O. McClellan

Received: 15 March 2011 / Accepted: 22 May 2011 / Published online: 1 June 2011

Abstract The Clean Air Act (CAA) requires listing as criteria air pollutants those pollutants that arise from multiple sources and are found across the United States. The original list included carbon monoxide, nitrogen oxides, sulfur oxides, particulate matter, photochemical oxidants (later regulated as ozone), and hydrocarbons. Later, the listing of hydrocarbons was revoked and lead was listed. The CAA requires the EPA Administrator to set National Ambient Air Quality Standards (NAAQS) for these pollutants using the "latest scientific knowledge" at levels that, in the judgment of the Administrator, are "requisite to protect public health" while "allowing an adequate margin of safety" without considering the cost of implementing the NAAQS. The NAAQS are set using scientific knowledge to inform the Administrator's policy judgments on each NAAQS. Recently, there has been increasing tension and debate over the role of scientific knowledge versus policy judgment in the setting of NAAQS. This paper reviews key elements of this debate drawing on the opinion of Supreme Court Justice Stephen Breyer, in *Whitman v. American Trucking Associations*, to resolve the conundrum posed by the CAA language. I conclude that scientists should carefully distinguish between their interpretations of scientific knowledge on specific pollutants and their personal preferences as to a given policy outcome (i.e., specific level and form of the NAAQS), recognizing that these are policy judgments as to acceptable levels of risk if the science does not identify a threshold level below which there are no identifiable health risks. These policy judgments are exclusively delegated by the CAA to the EPA Administrator who needs to articulate the basis for their policy judgments on the level and form of the NAAQS and associated level of acceptable risk.

Keywords Clean Air Act · Criteria pollutants · Ozone · Particulate matter · Policy · Risk · Regulations

Introduction

In this paper, I briefly review key aspects of the Clean Air Act (1970) with regard to the setting of National Ambient Air Quality Standards (NAAQS) for criteria pollutants noting various landmark decisions. I address the primary or health-based Standards and do not consider the secondary or welfare-based Standards, although the core concepts are also relevant to the setting of the secondary Standards. I highlight actions of the last two EPA Administrators (Stephen Johnson and Lisa Jackson) and the Clean Air Scientific Advisory Committee (CASAC) related to the setting of NAAQS for particulate matter and ozone that serve to illustrate the growing tension and debate over the role of scientific knowledge and policy judgments in the setting of NAAQS. I conclude with recommendations for the role of CASAC in synthesizing and interpreting the science on criteria pollutants and offering scientific advice that informs the EPA Administrator's policy judgments on acceptable health risks that, in turn, are linked to the level and statistical form of the NAAQS primary Standard.

This paper was presented in the concluding plenary session on "Regulatory and Policy Implications" at the "American Association for Aerosol Research International Specialty Conference: Air Pollution and Health: Bridging the Gap from Sources to Health Outcomes," March 22–26, 2010, San Diego, CA.

R. O. McClellan (✉)
Toxicology and Human Health Risk Analysis,
13701 Quaking Aspen Pl NE,
Albuquerque, NM 87111, USA
e-mail: roger.o.mcclellan@att.net

The Clean Air Act

The Clean Air Act (CAA), initially passed in 1963, is the principal national statute in the United States concerned with air quality. The original CAA (1963) directed the then Department of Health, Education and Welfare (HEW) to prepare, "compile and publish criteria on the effects of air pollutants," hence the identification of "criteria pollutants" and "criteria documents" summarizing the scientific knowledge on certain air pollutants arising from multiple sources and found across the United States as a basis for Standard setting. The National Air Pollution Control Administration (NAPCA) within HEW was assigned responsibility for administering the CAA. When the U.S. Environmental Protection Agency (EPA) was created in 1970, responsibility for administering the CAA was transferred from NAPCA to the new agency. Bachmann (2007) provides an in-depth review of the evolution of Air Quality Management in the United States from 1900 through 2006, with emphasis on the NAAQS, for those readers interested in an in-depth coverage of the topic. John Bachmann prepared his historical review soon after he retired from EPA's Office of Air Quality Planning and Standards where he had a central role for more than three decades in the setting of NAAQS for all the criteria pollutants. Readers interested in legal details of the CAA will find the summary of Martineau and Novello (2004) useful.

In 1970, amendments to the CAA (1970) were passed that required the listing of air pollutants that "may reasonably be anticipated to endanger public health and welfare" and to issue air quality criteria for them. These air quality criteria are to "accurately reflect the latest scientific knowledge useful in indicating the kind and extent of all identifiable effects on public health or welfare which may be expected from the presence of [a] pollutant in the ambient air, in varying quantities."

The pollutants originally designated as "criteria pollutants" because of their ubiquitous distribution and potential to endanger health were photochemical oxidants (later regulated as ozone), particulate matter (later regulated as total suspended particulates, then as PM_{10}, and $PM_{2.5}$), carbon monoxide, sulfur oxides (regulated as sulfur dioxide), nitrogen oxides (regulated as NO_2), and non-methane hydrocarbons (later dropped as a criteria pollutant). The EPA (1971) established NAAQS for these pollutants, soon after the Agency was created, using existing scientific documentation, i.e., criteria. As I will discuss below, the EPA later added lead as a criteria pollutant with legal prodding from the National Resources Defense Council.

In 1977, several key amendments were made to the CAA (1977). Concern about slow action of the EPA in preparing criteria documents and reassessing NAAQSs prompted a legislated requirement that the NAAQSs be reevaluated not later than January 1, 1980, and at 5-year intervals thereafter. Reevaluation was not intended to automatically result in changes in the NAAQSs for a pollutant; rather, reevaluation was intended to ensure that the scientific database was reviewed and that the NAAQSs were consistent with current knowledge. To my knowledge, this requirement for mandatory review every 5 years is unique to the setting of the NAAQS in the United States. Indeed, I know of no other statute calling for an updating of the science and reconsideration of the Standard every 5 years.

Peer review of the earliest criteria documents prepared by the EPA was carried out by various committees of the agency's Science Advisory Board as I will discuss later. A 1977 amendment to the CAA institutionalized the peer-review process for the NAAQS (CAA 1977). The amendment requires the EPA Administrator to appoint an independent scientific committee, composed of seven members, including at least one member of the National Academy of Sciences, one physician, and one person representing state air pollution control agencies to advise the Administrator on the science informing the policy judgments made in setting the NAAQS. The EPA has implemented this provision of the CAA by appointing a Committee, which designated itself as the CASAC. The CASAC is directly responsible to the EPA Administrator, although it functions administratively as one of the standing committees of the EPA Science Advisory Board. Traditionally, the requirement for one CASAC member to be a member of the National Academy of Sciences has been broadly interpreted to also include membership in either the National Academy of Engineering or the Institute of Medicine. To complement the expertise of regular members of the CASAC, consultants with specialized expertise usually have been added to the review panels for specific pollutants.

The CAA was amended again in 1990 (CAA 1990). Although major changes were made in the CAA with these amendments, especially with regard to the regulation of hazardous air pollutants, there were no changes in the fundamental approach to dealing with the setting of NAAQS for criteria pollutants. However, there were changes in the CAA that have had major impact on the regulation of emissions of PM and precursors especially from large power plants.

National Ambient Air Quality Standards

Section 109 of the CAA (1970) directs the Administrator to propose and promulgate "primary" and "secondary" NAAQSs for criteria pollutants identified under Section 108. The primary Standards are to be set to protect public health; secondary Standards are to be set to protect the

public welfare such as effects on soils, water, crops, visibility, and deterioration of property. In this paper, I focus on the use of scientific knowledge and judgment in the setting of the primary Standards. However, the issues discussed are also broadly applicable to the setting of secondary Standards.

Section 109(b)(1) defines a primary NAAQS as one that "the attainment and maintenance of which in the judgment of the Administrator, based on the criteria and allowing an adequate margin of safety, is requisite to protect the public health." The margin of safety, as interpreted by the EPA, is intended to address uncertainties associated with inconclusive scientific and technical information at the time the Standard is set and to account for hazards that research has not yet identified.

The primary Standards are intended to protect against "adverse effects, not necessarily against all identifiable effects of changes produced by a pollutants." Although Congress did not rigorously define an adverse effect, it did provide general guidance in the legislative history of the debate on the CAA (Library of Congress 1974). Congress was concerned with effects ranging from cancer, metabolic and respiratory disease, and impairment of mental processes to headaches, dizziness, and nausea.

Congress also noted concern for sensitive population groups in setting the NAAQSs. In particular, Congress noted that the Standards should protect "particularly sensitive citizens such as bronchial asthmatics and those with emphysema who in the normal course of daily activity are exposed to the ambient environment." This has been interpreted to exclude individuals who are not performing normal activities, such as individuals who are hospitalized. Further guidance was given noting that the Standard is statutorially sufficient whenever there is "an absence of adverse effect on the health of a statistically related sample of persons in sensitive groups from exposure to the ambient air."

The challenge of interpreting the language of the CAA was noted in an editorial by Donald Kennedy on "Risk versus Risk" published when he served as Editor-in-Chief of Science (Kennedy 2005). He wrote—"In the United States and some other industrial democracies, where people and their governments tend to be risk averse, legislatures, courts, and administrative entities usually create a presumption favoring more safety rather than less. The definitions of risk in law are often vague ("reasonable certainty of no harm" or "adequate margin of safety") and are likely to encourage an unrealistic belief that risks can be minimized or even eliminated altogether." I think Kennedy has captured the conundrum posed by the language of the CAA, a conundrum that has been addressed by Supreme Court Justice Stephen Breyer as I will relate later.

Standard-setting process

The process for developing and issuing NAAQS is quite complex. Key elements of the process, as used until quite recently, include preparation and review of (a) criteria document, (b) staff paper, (c) more recently a risk assessment, and (d) a regulatory decision package leading to the Administrator's policy judgment decisions as to the proposed and final NAAQS which are published in the Federal Register. Traditionally, CASAC focused its attention on reviewing the Criteria Documents and Staff Papers and, more recently, a formal Risk Assessment. As an aside, the process was changed at the end of 2006 (Peacock 2006) with an Integrated Science Assessment and Policy Assessment Document replacing the Criteria Document and Staff Paper. Time will tell if these changes really improve the overall process.

In addition to the documents noted above, the Agency now prepares a Regulatory Impact Analysis which is required under Executive Order 12866 issued by President Clinton (1993) that applies to economically significant rules that have "an annual effect on the economy of $100 million or more or adversely effect in a material way the economy, a sector of the economy, productivity, competition, jobs, the environment, public health or safety or site, local, or tribal governments or communities." The Regulatory Impact Analysis is not considered during the NAAQS rulemaking process given the prohibition of consideration of cost in the setting of the NAAQS, as will be discussed later.

The first Criteria Document prepared and released by the EPA addressed lead as a criteria air pollutant. This document was prepared and the review initiated before a Clean Air Scientific Advisory Committee was mandated by the CAA Amendments of 1977. Lead was not one of the original criteria pollutants. In 1975, the Natural Resources Defense Council (NRDC), with legal leadership from Attorney David Schoenbrod, sued EPA to have lead listed as a criteria pollutant. The EPA argued that it was already dealing effectively with reducing lead in air through its program to remove lead from gasoline. The Second Circuit Court disagreed (NRDC v. Train 1976) and on March 1, 1976, ordered EPA to identify lead as a criteria pollutant and begin the process of developing a NAAQS. At the time, EPA's Science Advisory Board (SAB) was in the process of assuming review responsibility for scientific activities across the Agency consolidating review functions brought to EPA from its predecessor organizations such as the National Air Pollution Control Administration (NAPCA). The EPA had just disbanded the National Air Quality Criteria Advisory Committee which had operated under NAPCA as well as other media specific advisory committees in favor of a series of discipline-oriented Committees; e.g., health, engineering and ecology.

In 1976, I was asked, as a member of the SAB Executive Committee, to chair an ad hoc Committee to review the criteria document on lead. Preparation of this document had already been initiated by EPA in anticipation of the Second Circuit Court decision. It was prepared by a Criteria and Special Studies Office within the Office of Research and Development located at EPA's Health Effects Laboratory in Research Triangle Park, NC. The first draft, released November 18, 1976, was viewed as unacceptable by the Ad Hoc Committee. The Committee was concerned with the poor scientific quality of the document. In addition, as noted by Bachmann (2007), the Committee was concerned that the document recommended a specific numerical Standard, a value of 5 $\mu g/m^3$, which was inconsistent with the intent of the CAA to separate the scientific assessment of the relevant criteria and the setting of the specific NAAQS.

The views of the Ad Hoc Committee members varied. Indeed, some members wanted the Committee to assume responsibility for re-writing the Criteria Document and recommending a specific Standard. As Chair, I emphasized our role was advisory to the Administrator, not to serve as substitutes for EPA staff to prepare the Criteria Document. The EPA proceeded to prepare a second draft which was released on May 27, 1977. The Committee viewed it as improved, but felt it was still not adequate for setting a lead Standard. The Agency proceeded to develop a third draft released on August 22, 1977. The Committee offered modest comments on the third draft which were considered by the Agency as it prepared the final criteria document released on December 14, 1977 (EPA 1977a) which served as a basis for the proposed lead NAAQS (EPA 1977b). As Chair, I conveyed to the Agency the view that the final version—"accurately reflected the available scientific literature and provided an adequate scientific basis for promulgation and issuance of a Standard for airborne lead." The first lead NAAQS was issued in 1978 (EPA 1978).

The experience with the lead criteria document served as a stimulus for EPA to create a separate Environmental Criteria Assessment Office within the Agency's Office of Research and Development. For three decades, this office was headed by Lester Grant. Grant originally came to the EPA from the University of North Carolina-Chapel Hill as an Inter-Government Personnel Act assignee to assist with revision of the criteria document on lead.

As noted by Bachmann (2007), the Office of Air Quality Planning and Standards (OAQPS) prepared an analysis to support the Lead Standard which was reviewed by EPA scientists, policymakers and the public. However, it was not reviewed by the SAB Ad Hoc Committee. That analysis served as a basis for the proposed NAAQS for lead (EPA 1977b) and the final lead NAAQS (EPA 1978). Bachmann (2007) has noted—"As for all NAAQS decisions, the final choice on the Standard was constrained and informed by the scientific information, but ultimately based on the policy judgment of a politically responsible decision-maker, the EPA Administrator. After consideration of and reaction to public comments, and review and discussion on the final package by OMB, the Administrator promulgated a Pb Standard of 1.5 $\mu g/m^3$ quarterly average in TSP." I strongly agree with Bachmann's first sentence assessment of the role of scientific information informing the policy judgments of the EPA Administrator. This will be a recurring theme in the remainder of this paper.

In many ways, the experience EPA gained in setting the lead NAAQS influenced the NAAQS process for subsequent NAAQS decisions. The OAQPS analysis evolved into preparation of formal Staff Papers that would be subjected to review by the CASAC. The first activity of the newly created CASAC, initially chaired by Sheldon Friedlander, was the review of a combined criteria document for particulate matter and Sulfur Oxides. Subsequently, separate addenda were prepared for Sulfur Oxides and particulate matter and separate Standards issued for the two pollutants. Sulfur Dioxide was identified as the indicator for Sulfur Oxides and Total Suspended Particulate (TSP) as the indicator for particulate matter.

Without going into the administrative or legal details, it is important to note that EPA, in carrying out mandated NAAQS actions in the early days, used an "informal rulemaking process" to propose and promulgate Standards (Bachmann 2007). The informal process focused on the end product, the NAAQS. The process was not always well documented as to how decisions were reached on the four elements of each NAAQS; the indicator, averaging time, specific numerical concentration and the statistical form. The DC Circuit Court of Appeals subsequently found that the record of this informal process did not give the Court a sufficient basis to complete its judicial review of the rules that were promulgated. This led to the final rule for the secondary Sulfur Dioxide Standard being revoked in 1973 as recounted by Berry (1984) in his review of NAAQS decision-making. This judicial decision led EPA to develop more rigorous procedures, including documentation, for the setting of each NAAQS (Pedersen 1975). As noted by Bachmann (2007), these procedures addressed the following points: "(1) EPA was to make available to the public the information and technical methodologies it relied upon by the time of proposal; (2) the preambles to proposal and final rules were to provide a detailed explanation of EPA's decision; (3) EPA was required to respond to all "significant" comments on the proposal by the time it issues its final rule; and (4) all of the above documents, analyses, preambles, and responses constituted the record that the court would examine in reviewing the final Standard decision. Objections not raised

in the record could not be raised in court. The halcyon days of a speedy NAAQS process were over." I agree that the speed of the process was reduced, however, I would add that the transparency of the process was also substantially improved. Congress apparently agreed and these provisions were substantially codified by the CAA Amendments of 1977.

EPA's implementation of the CAA, especially its setting of NAAQS even with improved documentation, has been a matter of continuing controversy and litigation (some persons might argue that controversy and litigation were enhanced by improved documentation in the record). Bachmann (2007) summarizes many of the key legal cases in his review. In this paper, I will only highlight certain of the key legal cases.

The 1997 revisions of the Ozone NAAQS (EPA 1997a) and Particulate Matter NAAQS (EPA 1997b) proved to be very contentious, including the discussions within CASAC. The CASAC PM Panel members had a range of views on the $PM_{2.5}$ Standard that was being set for the first time supplementing the PM_{10} Standard. This range of views was clearly articulated in the CASAC Chair's letter (Wolff 1996) to the Administrator by including a Table showing the views of each individual.

The contentious nature of the debate over these revised NAAQS prompted Administrator Browner to involve President Clinton. Bachmann (2007) recounts that Administrator Browner had a 1-h meeting on these Standards with the President—"she reported that the President quickly accepted her decision and spent much of the time discussing how to reduce unnecessary burdens in the implementation process. This resulted in some of us writing the first draft of a letter that was later sent by President Clinton (Clinton 1997) to EPA directing implementation be carried out so as to "maximize common sense, flexibility, and cost effectiveness."" Not surprisingly, President Clinton (New York 1997) had a role in announcing the tighter Standards which included for the first time a separate $PM_{2.5}$ Standard to supplement the PM_{10} Standard and a shift from a 1-h averaging time to an 8-h averaging time Standard for Ozone.

The issuance of a revised PM NAAQS triggered the case of *American Trucking Associations v. EPA* (ATA 1999). The Court found "the growing empirical evidence demonstrating a relationship between fine particle pollution and adverse health effects amply justifies establishment of new fine particulate Standards." The Court went on to find "ample support" for EPA's decision to regulate coarse particulate pollution, but vacated the 1997 PM_{10} Standards, concluding in part that PM_{10} is a "poorly matched indicator for coarse particulate pollution" because it includes fine particulates which were separately regulated as $PM_{2.5}$. Subsequently, EPA removed the vacated 1997 PM_{10} Standard allowing the 1987 PM_{10} Standard to remain in place along with the new $PM_{2.5}$.

In addition, the three judge panel held, two to one, that EPA's approach to setting the level of the PM and Ozone Standards in 1997 effected "an unconstitutional delegation of legislative authority." The Judicial Panel found that "the factors EPA uses in determining the degree of public health concern associated with different levels of ozone and particulate matter are reasonable." However, it remanded the rule to EPA. The Judicial Panel stated that when the Agency considers these factors for potential non-threshold pollutants "what EPA lacks is any determinate criterion for drawing lines" to determine the level at which the Standards should be set. The Judicial Panel also found that the Administrator, under the CAA, is not permitted to consider the cost of implementing these Standards in setting them.

Not surprisingly, the nature of the Circuit Court opinion resulted in cross appeals being filed on the several issues. The Supreme Court in February 2001 issued a unanimous opinion upholding EPA's position on both the Constitutional and cost issues (*Whitman v. American Trucking Associations* 2001). On the Constitutional issue, the Supreme Court held that the statutory requirement that the NAAQS be "requisite" to protect public health with an adequate margin of safety sufficiently guided EPA's discretion, affirming EPA's approach of setting Standards that are neither more nor less stringent than necessary.

Supreme Court Justice Breyer, who participated in the *Whitman v. American Trucking Associations* Case, is well known and highly regarded for his opinions and writings on risk assessment and regulation (Breyer 1982, 1993). Thus, it is not surprising that he took the opportunity in *Whitman v. American Trucking Associations* (2001) to offer comments on the Standard-setting process and, specifically, the identification of the level of the NAAQS and the associated level of health risk. While concurring that EPA cannot consider the costs of implementing the NAAQS, he went on to note—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.

The 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."

The case of *Whitman v. American Trucking Associations* (2001) is widely cited for the conclusion that EPA cannot consider the economic costs of compliance in the setting of NAAQS. Unfortunately, in my opinion, insufficient attention is given to the thoughtful guidance of Justice Breyer on exercising policy judgment in deciding on an acceptable level of health risk, a judgment that in turn determines the level and statistical form of each NAAQS. It is interesting that Justice Breyer's opinion appeared in Administrator Johnson's notice of the Ozone NAAQS (EPA 2008), but did not appear in Administrator Jackson's "reconsideration" proposal for ozone (EPA 2010a) which will be discussed later.

Paradigm shift

At this juncture, it is appropriate to note that it is my view that a paradigm shift has taken place in the use of scientific knowledge and policy judgments in the selection of the level and form of each NAAQS over the past four decades. In my opinion, the paradigm shift has been driven in part by the nature of the growing body of scientific evidence of pollution effects. In the 1970s, most scientists and regulators viewed the criteria pollutants as having a threshold in the concentration–response relationship for non-cancer endpoints, the major concern for the criteria pollutants. This was different than the prevailing view for cancer causing agents which were assumed to have linear, non-threshold, concentration–response relationships.

In the early 1970s, the available data on each criteria pollutant were quite modest, with attention in the review process focusing on only a few epidemiological studies. For those few studies, attention often focused on whether a relative risk on the order of 2.0 was observed and whether it was statistically significant or not. For a given criteria pollutant there were few, if any, controlled human exposure studies. The data from laboratory animal studies had frequently been acquired in short-term studies with exposure concentrations much higher than ambient concentrations. This raised questions about extrapolation from laboratory animals to humans and high to low exposure concentrations. The general approach taken to evaluating the published studies was to identify the lowest levels where effects were statistically significant and assume this was the inflection point in the concentration–response relationship. It could then be readily argued that setting the Standard at a lower concentration than that at which

effects were observed satisfied the requirement for "an adequate margin of safety."

In contrast, the most recent reviews of the criteria pollutants have involved thousands of papers with observations ranging from the human population level to studies of intact laboratory animals to studies of effects of air pollutants on cells and molecules. Despite the huge number of published studies, the focus has ultimately centered in the Staff Paper on the results of a few studies where attention turns to the relevance of the results for informing policy judgments on the level and statistical form of the Standard. For the epidemiological studies, the debate often focuses on whether relative risks of less than 1.1 for excess morbidity and mortality are significant. Of course, the specific relative risk number is dependent on the denominator being used. For controlled exposure clinical studies, attention has focused on the lowest levels with statistically significant changes and whether the changes are adverse.

A news report (Taube 1995) in Science, that I view as a classic report, highlighted the issues involved in the search for subtle links between diet, lifestyle, or environmental factors and disease, especially using retrospective observational studies. I especially liked the quote at the end attributed to UCLA Professor Greenland in offering advice to his "most sensible, level-headed, estimatable colleagues." Remember, he says—"there is nothing sinful about going out and getting evidence, like asking people how much do you drink and checking breast cancer records. There's nothing sinful about seeing if that evidence correlates. There's nothing sinful about checking for confounding variables. The sin comes in believing a casual hypothesis is true because your study came up with a positive result, or believing the opposite because your study was negative."

It is interesting to note that CASAC discussions of criteria pollutant effects have frequently focused initially on the level of the Standard, devoid of any consideration of the statistical form of the level. This approach was in keeping with traditional practice in the setting of Standards such as Threshold Limit Values for occupational exposures to chemicals (McClellan 1999, 2010c). That approach has traditionally involved a review of the available human data on a toxic chemical to determine a no-observed effect level, or the lowest observed effect level, and then use of a safety factor to arrive at an acceptable exposure level set at a lower level. In the absence of adequate human data, laboratory animal data are used and an additional safety factor applied to account for the potential that the animal observations might not adequately predict human effects. This approach was routinely used for a wide range of health responses that were assumed to have an exposure-response relationship that exhibited either a true or practical threshold, an excess of effects above some level and an absence of effects below that level. A review of the earliest Criteria Documents and, indeed, also the Staff Papers, documents that a similar line of reasoning was used in the setting of the NAAQS—identify levels where an increase in effects is observed and then set the Standard at a lower level.

The implementation of Standards set with this approach soon revealed that if the Standard was to be rigorously enforced, i.e., no exceedances of the specific level of the Standard, the practical effect would be to cause average levels of the pollutant to be reduced to levels far below the Standard so as to avoid the occasional high concentration exceeding the Standard. Fortunately, common sense prevailed and the EPA, over time, moved to the practice of routinely linking attainment of the specific level of the Standard to a statistical form such as the 98th percentile 24-h concentration averaged over 3 years, or the fourth highest 8-h average concentration during a 3-year period. In my experience, most of the attention of the CASAC in the NAAQS-setting process has focused on the level of the Standard with limited discussion of the statistical form of the Standard. In doing so, there has been a failure to recognize that the stringency of the Standard and the degree of health protection provided depends on both the level and statistical form of the Standard for a particular indicator and averaging time. In fact, there have been occasions when CASAC has deliberated at length on the level of a prospective Standard and, then in a casual manner, turned its attention to what would be the appropriate statistical form for that level. That this is the case is not surprising since few scientific papers discuss the implications of the reported results in terms of the frequency with which a given health effect may be observed.

The challenges of selecting appropriate averaging times and statistical forms for the NAAQS are substantial. The original epidemiological and toxicological studies that provide the scientific information that should inform the setting of the NAAQS do not always report results with an averaging time that is the same as used for the Standard. Hence, the need to make extrapolations from results reported based on one metric, such as average daily exposure, to second metric, such as an 8-h or shorter averaging time. The setting of Standards at extreme values, the 98th percentile for NO_2 (EPA 2010b) and the 99th percentile form as done with the 1-h averaging time Standard for SO_2 (EPA 2010c), results in extremely stringent Standards that at best are only very loosely related to the underlying data.

In my view, decisions on the selection of specific levels and averaging times for the NAAQS are policy judgments properly reserved to the Administrator informed by the available scientific knowledge. In the 1990s, concurrent with the increasingly widespread use of formal risk analysis

procedures across society (McClellan 1999, 2010c), EPA moved to quantify the health benefits associated with setting the NAAQS at various levels, with an associated statistical form. I must admit to being an early advocate of formal quantification of health benefits of various levels and forms of the prospective Standards. I viewed the approach then and I still do today, as a way to synthesize the science so it could provide useful guidance to the Administrator for making policy decisions. I did not envision that some advocates of quantitative risk analysis would actually view the results of the analyses as being highly accurate projections of potential health benefits expressed to two or more significant figures, sometimes without any indication of uncertainty.

The quantification of health effects potentially associated with various levels and forms of the Standards requires several kinds of input. First and foremost, it requires some knowledge of the nature of the concentration–response relationships for various temporal metrics for the pollutant in question. Typically, the response term is expressed as excess risk per unit of increased concentration over some range of ambient concentrations. The question then becomes one of whether the relationship is linear and whether there is a threshold level below which the coefficient for excess risk does or does not hold. The issue of whether there are or are not thresholds for non-cancer health endpoints is very contentious and a subject of on-going debate (White et al. 2009; Rhomberg et al. 2011). A related issue becomes the selection of suitable reference baseline statistics for the particular health effects. An additional question becomes the appropriate population to be evaluated—a single city, multiple cities or the population of the United States. It is obvious that there are substantial uncertainties associated with each component of the analyses.

With the use of linear, no-threshold, concentration–response models, the EPA has on some occasions calculated estimated excess morbidity and mortality effects attributable to the specific pollutant down to background concentrations. The Health Risk Assessment (EPA 2007b) and the Regulatory Impact Analysis (EPA 2007d) for the 2008 Ozone NAAQS serve as examples. Further, dependent on the assumptions made with regard to how ambient concentrations of the pollutant would change in response to various levels and forms of the Standard, estimated health effects avoided (i.e. health benefit) may be calculated. A key consideration as to whether these benefits can be realized relates to whether the roll-back in air concentrations that is assumed in the analysis as a result of implementation of the new Standard can actually be realized. In part, the validity of the analyses relate to how realistic the assumptions have been with regard to background levels. A discussion of this issue for ozone can be found in McClellan et al. (2009). Indeed, as the levels of the Standards are ratcheted down toward background levels, there is increasing uncertainty as to whether there are any health effects attributable to single pollutants and even greater uncertainty as to the magnitude of the health benefits associated with any new lower Standard. The use of single pollutant models for estimating benefits also raises the issue of double-counting of benefits as the benefits of the individual pollutants are aggregated.

Hence, the paradigm shift. It is apparent that in setting the earliest NAAQS some individuals, including CASAC members, envisioned that the Standards were being set at levels protective of public health with an adequate margin of safety based on threshold concentration–response models. In short, if there were health effects at the level and form of the selected NAAQS, they were viewed as de minimis. In contrast, more recent NAAQS have been set at levels which the CASAC and EPA characterize as having residual health effects even if the Standard were to be attained. The central question remains—how low is low enough? I view the answer as a policy judgment informed by science that can only be made by the EPA Administrator.

Recent action on revision of the particulate matter and ozone NAAQS

It is instructive to now turn our attention to the most recent actions of EPA with regard to the revision of the PM NAAQS in 2006 (EPA 2006b), the revision of the Ozone NAAQS in 2008 (EPA 2008) and the "reconsideration" proposal (EPA 2010a) for a further revision of the Ozone NAAQS in 2011

In the initial discussion, I will focus on the EPA's 2006 revision of the $PM_{2.5}$ Standard. The science that informed the setting of that Standard was summarized in a Criteria Document (EPA 2004). This, in turn, provided the basis for the Staff Paper (EPA 2005). The central issue was the level and associated form of the two different averaging time Standards, a 24-h averaging time and an annual Standard. The first Standards using $PM_{2.5}$ as an indicator were set in 1997 (EPA 1997a). The 24-h averaging time Standard was set at 65 $\mu g/m^3$. The 24-h $PM_{2.5}$ Standard of 65 $\mu g/m^3$ was attained when the 3-year average of the 98th percentile of the concentrations at each population-oriented monitor was not exceeded. The Annual Standard was set at an annual arithmetic mean of 15 $\mu g/m^3$. The annual Standard was attained when the 3-year average of the weighted $PM_{2.5}$ concentration from single or multiple community-oriented monitors did not exceed 15 $\mu g/m^3$. Recall that the 1997 $PM_{2.5}$ Standard was originally intended to supplement and, in part, replace the PM_{10} (annual arithmetic mean of 50 $\mu g/m^3$ and 24 h average of 150 $\mu g/m^3$) Standard set in 1987. That PM_{10} Standard had replaced the earlier Total Suspended Particulate Standard promulgated in 1971.

I participated as a member of the CASAC Panel that provided advice on the setting of the $PM_{2.5}$ Standard in 1997. There was much discussion about the uncertainty associated with the shift from a PM_{10} to a $PM_{2.5}$ Standard, especially the uncertainty in a shift from dependence on only the PM_{10} indicator to $PM_{2.5}$ indicator. There was strong scientific support for introducing the $PM_{2.5}$ indicator, although at the time, there was limited epidemiological data from studies in which $PM_{2.5}$ had actually been measured. There was no clear scientific evidence on the presence or absence of a threshold in the concentration–response relationship for either acute or chronic responses. The big issues related to the levels and associated form—"how low was low enough?" The prevailing tone in hallway conversations focused on two points. First, it was argued that it was important to introduce a $PM_{2.5}$ indicator which, in turn, would mandate the monitoring of $PM_{2.5}$. The availability of the $PM_{2.5}$ monitoring data would then allow the conduct of epidemiological studies to directly evaluate a potential concentration–response association for this indicator. Second, it was argued that in the absence of convincing data on $PM_{2.5}$ the final action contemplated by the Agency should not represent a drastic increase in the stringency of the PM Standard. In my opinion the new $PM_{2.5}$ annual Standard set at 15 $\mu g/m^3$ *did* increase the stringency of the PM Standard and represented a policy judgment call on the part of the Administrator that was very precautionary. In contrast, in my opinion, the setting of $PM_{2.5}$ 24-h averaging time Standard at 65 $\mu g/m^3$ was much less precautionary. The level and form of the new Standards was as follows:

(1) The annual $PM_{2.5}$ Standard is met when the 3-year average of the annual arithmetic mean $PM_{2.5}$ concentrations, from single or multiple community-oriented monitors, is less than or equal to 15 $\mu g/m^3$, with fractional parts of 0.05 or greater rounded up.
(2) The 24-h $PM_{2.5}$ Standard is met when the 3-year average of the 98th percentile of 24-h $PM_{2.5}$ concentrations at each population-oriented monitor within an area is less than or equal to 65 $\mu g/m^3$, with fractional parts of 0.5 or greater rounded up.
(3) The form of the previous 24-h PM_{10} Standard is revised to be based on the 3-year average of the 99th percentile of 24-h PM_{10} concentrations at each monitor within an area.

Review of the PM Standard that would lead to revision of the 1997 PM Standard moved forward in the early 2000s. In 2004, as the new Criteria Document for PM was reviewed, it was decided that the CASAC would abandon CASAC's practice of issuing "closure letters." "Closure Letters" had traditionally been sent by the CASAC Chair to the EPA Administrator at key junctures, such as completion of revision of a Criteria Document or Staff Paper, signifying the work product was scientifically acceptable for regulatory decision-making. Some individuals had viewed the "closure letters" as a way by which CASAC impeded progress in the setting of NAAQS in a timely manner. I viewed the "closure letters" as an effective approach to ensuring that EPA was preparing documents that included the latest scientific information and analyses, even if it required the Agency to develop Revisions or Addendums.

After reviewing and commenting on the Criteria Document (EPA 2004) and Staff Paper (EPA 2005), CASAC recommended that the 24-h $PM_{2.5}$ Standard be set in the range of 25–35 $\mu g/m^3$ and the annual $PM_{2.5}$ Standard be set in the range of 13–14 $\mu g/m^3$ (Henderson 2005, 2006a; Table 1).

There was strong pressure within the CASAC PM Panel to provide consensus advice to the Administrator. In the end, two consultant members of the PM Panel who had both served as Chair of CASAC (myself and another) did not deem it appropriate to join with other members of the Panel in endorsing the specific levels others wished to recommend to the Administrator. I held strongly to the view that the difference between leaving the Standard at 15 $\mu g/m^3$ and reducing it to 14 $\mu g/m^3$ was not a scientific decision, but rather a matter of policy judgment that should be left to the discretion of the Administrator. In my opinion, Administrator Johnson, as the politically responsible decision-maker (using the words of John Bachmann 2007 in describing the 1974 Lead NAAQS decision) was not bound by the recommendations of CASAC as they were an advisory committee. In my opinion, the Administrator alone had the authority to make policy judgment calls in retaining or revising the annual $PM_{2.5}$ Standard, then at 15 $\mu g/m^3$ and the 24-h $PM_{2.5}$ Standard, then at 65 $\mu g/m^3$ (EPA 1997a). The Administrator issued a final rule with the annual $PM_{2.5}$ Standard retained at 15 $\mu g/m^3$ and the 24-h Standard reduced to 35 $\mu g/m^3$ (EPA 2006b).

Table 1 National ambient air quality standards for $PM_{2.5}$ and ozone, the old standard, CASAC recommendations and administrator's final rule

Indicator (unit)	Old standard	CASAC	New standard
$PM_{2.5}$—24 h ($\mu g/m^3$)	65[a]	30–35[b]	35[c]
Annual ($\mu g/m^3$)	15[a]	13–14[b]	15[c]
Ozone—8 h (ppb)	84[d]	60–70[e]	75[f]

[a] EPA 1997a, b

[b] Henderson 2006a, b; Henderson et al. 2006c

[c] EPA 2006b

[d] EPA 1997a, b, set at 0.08 ppm which by rounding convention equals 84 ppb

[e] Henderson 2007, 2008

[f] EPA 2008

After the final PM rule was issued in 2006 (EPA 2006b), the seven formal members of CASAC (Henderson et al. 2006c) sent a letter to the Administrator expressing concern that the EPA Administrator had not decreased the $PM_{2.5}$ Annual Standard from 15 $\mu g/m^3$ to 13–14 $\mu g/m^3$ in combination with the setting of the 24-h Standard at 35 $\mu g/m^3$, the upper end of the ranges they had recommended. In my view, the CASAC recommendation that the Administrator had to reduce the annual Standard by at least 1 $\mu g/m^3$ indicated that the CASAC failed to appreciate that the setting of any NAAQS involves policy judgments, reserved by the CAA to the EPA Administrator, informed by the science. Presumably, the CASAC would have found it acceptable if the Administrator had reduced the Annual $PM_{2.5}$ Standard from 15 to 14 $\mu g/m^3$, or even to 13 $\mu g/m^3$.

Perhaps it would be useful for me to elaborate on why I think it is not appropriate for CASAC to recommend a bright line upper bound on the NAAQS, even assuming no change in the statistical form of the Standard. The Committee, when commenting on the science undergirding the Standard, had noted that it had not identified a threshold in the ambient exposure concentration–response relationship for $PM_{2.5}$. Consistent with this assessment of the science, the EPA in its Risk Assessment had used a linear exposure concentration–response model to estimate risk that would be avoided and risks that would remain if the Standards were set at various specific levels and with an assumed statistical form. There were estimated risks associated with retaining the Standard at 15 $\mu g/m^3$ and reducing it to 14 or 13 $\mu g/m^3$. By endorsing a level of 14 $\mu g/m^3$ for the annual Standard, the CASAC was indicating its support for setting the Standard at a particular level of estimated risk. In my opinion, a decision on acceptable risk (i.e., the residual risk level when the Standard is attained) is a policy decision left to the discretion of the EPA Administrator under the authority of the CAA. The Committee's blended scientific and policy judgment advice would have been clearer if they had stated their specific advice by indicating both the specific numerical level and the associated morbidity and mortality. Of course, the estimates of morbidity and mortality should have had an indication of the associated uncertainties.

Let us now turn to revision of the Ozone NAAQS. Final action on revision of the Ozone Standard set in 1997 (EPA 1997b) followed almost 2 years after the decision on the $PM_{2.5}$ Standard. The ozone review included a Criteria Document (EPA 2006c) which summarized publications through 2005. This document served as the basis for a subsequent staff paper (EPA 2007a) and risk assessment (EPA 2007b). Again, CASAC (Henderson 2006b, 2007) offered very prescriptive advice on the level of the Standard indicating that the level of the revised 8-h averaging time Standard should be lowered to no greater than 0.070 ppm down from the 1997 Standard of 0.08 ppm which by rounding convention was effectively 0.084 ppm. The 1997 Standard is met when the 4th highest 8-h average value over a 3-year period does not exceed 0.084 ppm (Table 1).

The CASAC letter on the Ozone Staff Paper (Henderson 2007) commented on policy relevant background (PRB) noting "the Final Ozone Staff Paper does not provide a sufficient base of evidence from the peer-reviewed literature to suggest that the current approach to determining a PRB is the best method to make this estimation." The letter concludes with the statement—"Thus, PRB is irrelevant to the discussion of where along the concentration–response function a NAAQS with an averaging time that provides enhanced public health protection should be." The CASAC apparently failed to appreciate that identification of scientifically valid levels for PRB for different sections of the country can have a profound influence on realizable public health benefits (see discussion in McClellan et al. 2009) and the calculated benefit and residual risks for various levels and forms of the Standards.

As the Agency's activities on revision of the Ozone NAAQS were proceeding, I participated in June 2007 with a small group of scientists at a meeting held in Rochester, NY to discuss critical considerations in evaluating scientific evidence of health effects of ambient ozone. The discussions at the Rochester Conference focused on the scientific interpretation of the 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 non-anthropogenic origin and issues involved with conducting formal risk assessment of the health impacts of current and prospective levels of ambient ozone. The participants, while offering comments on the science informing the revision of the Ozone NAAQS, did not feel it appropriate to offer policy judgments on the level and form of the Ozone NAAQS then under consideration. A report based on the Rochester Conference has been published (McClellan et al. 2009). The deliberations at the Rochester Conference were summarized and included with my comments (McClellan 2007) submitted to the EPA Ozone Docket on the proposed Ozone Standard (EPA 2007c).

Administrator Johnson, in March 2008 (EPA 2008), issued a final revised Standard for Ozone with the primary 8-h average Standard set at 75 ppb retaining the statistical form the same as the 1997 primary Standard—the Standard is attained when the fourth highest 8-h average value over a 3-year period does not exceed 75 ppb. The CASAC was displeased with the policy judgment of Administrator Johnson to set the Standard at 75 ppb rather than heeding

their recommendation to set the Standard in the range of 0.060–0.070 ppm (Henderson 2008). As an aside, Administrator Johnson also decided to set the secondary Standard for Ozone equal to the primary Standard. In doing so, he did not heed CASAC's advice to set a secondary Standard with a different cumulative form. The CASAC had recommended a sigmoidally weighted W126 index, accumulated over 12 "daylight" hours and over at least the three maximum ozone months of the summer growing season (Henderson 2008).

Some CASAC members have argued that by giving the EPA Administrator a range (0.060–0.070 ppm), the CASAC had not taken away the Administrator's discretion in making policy judgments on the level and form of the NAAQS. To the contrary, I argue that the upper value in the range is in effect a bright line that CASAC has indicated the Administrator should not go above based on the science. In short, under the new paradigm, CASAC has defined for the Administrator the upper level of excess risk that CASAC deems acceptable, even though they have not clearly identified the specific health risk level associated with the 0.070 ppm level.

I firmly believe that Administrator Johnson's decisions on both the primary and secondary ozone Standards were consistent with the legislative authority accorded the Administrator under the CAA. Much was made of the fact that in the setting of the Ozone Standards, discussions took place between White House staff and, perhaps then President Bush, as the Standard was finalized. This is hardly surprising. Recall Bachmann (2007) recounted the discussions between President Clinton and Administrator Browner in 1997 and the draft memo to EPA Administrator Browner prepared by EPA staff for ultimate issuance over the signature of President Clinton (Clinton 1997).

As soon as President Obama was sworn in on January 20, 2009, the then-White House Chief of Staff, Rahm Emanuel, issued a memorandum (Emanuel 2009) stating —"It is important that President Obama's appointees and designees have the opportunity to review and approve any new or pending regulations." The Emanuel memorandum then proceeded to outline explicit conditions for what qualified as new or pending regulations—for example, "all proposed or final regulations that have not been published in the Federal Register" and "consider extending for 60 days the effective date of regulations that have been published in the Federal Register but not yet taken effect." The revised NAAQS for ambient ozone, published in the Federal Register, March 12, 2008 (EPA 2008), could hardly be viewed as new or pending in January 2009. Indeed, in the fall of 2008, the EPA had already initiated action on the next review of the Ozone NAAQS (Martin 2008). In initiating the new review, it was noted that CASAC advice on the previous review of the Standard represented "a mixture of scientific and policy considerations." Nonetheless, EPA Administrator Lisa Jackson in late 2009, decided to proceed with "reconsideration" of the final Ozone NAAQS rule issued in March 2008 (EPA 2008). The decision to proceed with a "reconsideration" proposal was formally announced in the Federal Register in January 2010 (EPA 2010a). The "reconsideration" proposal noted—"With respect to CASAC's recommended range of standard levels, EPA observed that the basis for CASAC's recommendation appears to be a mixture of scientific and policy consideration."

Administrator Jackson has stated that the "reconsideration" rule will be based on the same record used to propose the 2008 Standard, essentially the scientific information available through late 2005 and included in the 2006 Criteria Document (EPA 2006a). Recall the earlier discussion of EPA moving to a formal rulemaking process at the insistence of the Court. The approach of using the "old scientific record" was apparently taken with a view that it offered a "fast track" to a revision of the Ozone Standard without creating a new record. The "reconsideration" proposal (EPA 2010a) states that consideration will be given to setting the primary Standard set in the range of 60 to 70 ppb. The announced date for release of the final "reconsideration" Standard has continually shifted from August 2010 to October 2010 to December 2010 to July 2011. In accord with the review plan laid out in October 2008, the EPA staff proceeded with preparation of the Integrated Science Assessment reviewing the new scientific information to be considered in the next 5-year review triggered by promulgation of the March 2008 Ozone NAAQS. Ironically, the Integrated Science Assessment, the document replacing the old criteria document, for ozone, was released on March 2, 2011 (EPA 2011a), all while EPA's reconsideration of the old record remains pending.

I offered comments (McClellan 2010a) on the appropriateness of the Administrator proceeding with a "reconsideration" Standard for ozone and offered comments (McClellan 2010b) to the EPA Ozone Docket on the specifics of the proposal. In my view, the proposal for the Administrator to reconsider a rulemaking, the setting of a NAAQS, formally completed 9 months earlier by the previous Administrator in another Administration is without precedent. It has the potential to serve as a bad precedent with every change in Presidential Administration triggering a review of actions completed by the previous Administration with a view to potentially reconsidering the rules. In short, the new Administrator is saying "if I had been in office before I was appointed, I would have made a different policy judgment call." Administrator Jackson's use of the CASAC position in 2008 to justify the "reconsideration" action, in my opinion, moves CASAC out of its scientific advisory role into a strategic, policy-driving

Standard-setting role. This is troubling since Administrator Johnson, in issuing the 2008 Standard, had noted (perhaps with trepidation) that the CASAC recommendation "appears to be a mixture of scientific and policy considerations," a view informed by EPA staff analysis (Martin 2008). I agree with the assessment that CASAC, in recommending specific levels, is on a path of mixing scientific interpretations with policy judgments.

Administrator Jackson, in early 2011 (EPA 2011b), called on the CASAC to offer further clarification of the views it expressed earlier. The specific advice being solicited by the Administrator from CASAC is detailed in a memorandum from Lydia Wegman, Office of Air Quality Planning and Standards to CASAC (Wegman 2011). Many of the questions appear to be directed at attempting to distinguish between CASAC's interpretation of the old science and the policy judgments that resulted in CASAC's 60–70 ppb recommended range for the Standard. It proved challenging for CASAC to address these questions based only on the "old record" of pre-2006 science while ignoring the new scientific information on ozone (Samet 2011).

The substantial new scientific information on ozone that has been published in the 5 years since the Criteria Document (EPA 2006c) was prepared is documented in the recently released Integrated Science Assessment (EPA 2011b). The current drama over the "reconsideration" ozone rule has the potential to damage the credibility of CASAC by drawing it more tightly into the "regulatory web of policy judgments" that are the exclusive dominion of the Administrator under the authority of the CAA. My advice (McClellan 2011) to the Administrator and CASAC was to withdraw the "reconsideration" proposal and ask CASAC to expeditiously proceed with review of the new science now available in the Integrated Science Assessment (EPA 2011a).

Call for sound science

Over the last several decades, there have been increasingly loud calls from multiple quarters for using "sound science" to make regulatory decisions such as the setting of NAAQS. The call has come from both Non-Government Organizations (NGOs) representing multiple sectors, from Industry and from the scientific community. In my opinion, all of these groups and the individuals within them have difficulty separating the science from their policy-driven preferred outcomes. As a scientist and as a citizen, I strongly support the use of all the available scientific information to inform public policy decisions. In general, I think the efforts of individuals and organizations to critically review and synthesize relevant scientific information for the various Agency rulemaking activities has had a positive impact. This includes the situations in which original scientific data files were made available (actions that I applaud) and re-analyses conducted. Indeed, I think more such analyses should be conducted, especially when the original data were acquired with public funding. By the same token, I would urge industry groups to make available to other investigators data acquired under industry sponsorship.

What I decry, however, is the desire by some to label certain reviews or analyses as either "acceptable" or "dead on arrival" based on the source of funding without regard to scientific quality of the review or analyses. Over my career, I have encountered exceptionally high-quality reviews and analyses performed by scientists in academic, industrial, and environmental organizations with sponsorship from government, NGOs, and industry. I have also noted some reviews and analyses from these same quarters that I thought were of inferior scientific quality. In my opinion, scientific quality and rigor is not defined by the source of funding for the work.

I have great concern that the advocates of "sound science," be it NGO, academics or industry, may have unrealistic expectations as to what "sound science" can deliver. Sound science does not in and of itself make for sound decisions. As I have noted in this paper, science alone cannot identify an acceptable level of health risk, since such levels inherently represent a policy judgment call. Sound science can only inform what are ultimately policy judgments or political decisions. This is especially the case for the setting of NAAQS, in the absence of a clearly defined threshold, which involve decisions as to acceptable health risks which are linked to the level (and form) of the Standard.

Setting NAAQS at acceptable levels of risk

Let us now return to the critical issue of "how low is low enough?" for setting a specific NAAQS. It is apparent that the body of science on any given criteria pollutant today is such that it is difficult to argue that the current Standards, if attained, would result in a world that is free of any risk of adverse effects from air pollution on the populations of the United States. As Justice Breyer wrote, we live in a world that is not free of all risk. I draw guidance from Justice Breyer's statement on his interpretation of the words of the CAA—"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." The "her" in Justice Breyer's opinion is a reference to past EPA Administrator Christine Whitman.

However, in my opinion, the discretion that Justice Breyer assigns to the EPA Administrator does not extend to

the CASAC, either as individuals or acting collectively. Each of the individuals serving on CASAC may be an extraordinarily competent scientist or engineer or have other specialized knowledge of air quality and its health and environmental effects. Because of this special expertise, these individuals have a special role in interpreting the scientific knowledge that the Administrator will use in making policy judgments on the level and form of the Standard recognizing that the level and form, in turn, determine the level of acceptable risk that it is estimated Society will bear for that specific pollutant.

As broadly knowledgeable health and environmental scientists, CASAC members are in a unique position to offer advice to the Administrator that will provide the "comparative health consequences" context that Justice Breyer has called for in his opinion. For example, it would be refreshing if CASAC members were to more broadly draw on their experience as health specialists. In doing so, when debate begins on the public health significance of an excess risk of 0.1 for some health endpoint per 10 ppb increase in ozone at 60, 70, or 80 ppb averaged over 8 h, they could offer comments on the multiple factors that influence the health risks for that endpoint. This discussion, in my opinion, should even be extended to recognize that complex factors such as the socio-economic status of individuals have a profound influence on health (Table 2; Steenland et al. 2004). I will readily admit that differences in air quality associated with socio-economic status may have a role in the differences reported by Steenland et al. (2004) and other investigators. However, that admission does not serve as a basis for not providing scientific context to decisions on "how low is low enough" in setting NAAQS.

I suspect that this was the kind of input Administrator Bill Ruckelshaus was seeking when he noted in 1983 that a decision on the PM Standard "could not be made solely on science, and asked if under the statute "is there room to consider other non-scientific factors in making the major social policy judgment of picking a precise number from a range of scientifically justified values" (Bachmann 2007). Justice Breyer has answered former Administrator Ruckelshaus' question in the affirmative. Indeed, Justice Breyer has recommended the use of comparative health consequences as a context for Standard setting. In doing so, he has indicated that the boundaries of the relevant science for setting a NAAQS are not restricted exclusively to the health effects of the specific pollutant under consideration. This common sense approach has not been evident in many of the recent CASAC deliberations or the policy judgments of the Administration.

Table 2 The impact of socio-economic status on mortality (Steenland et al. 2004)

Mortality	Men	Women
All causes	2.02 (1.95–209)[a]	1.29 (1.25–1.32)
Heart disease	1.88 (1.83–1.93)	1.84 (1.76–1.93)
Stroke	2.25 (2.14–2.37)	1.53 (1.44–1.62)
Diabetes	2.19 (2.07–2.32)	1.85 (1.72–2.00)
COPD	3.59 (3.35–3.83)	2.09 (1.91–2.30)
Lung cancer	2.15 (2.07–2.23)	1.31 (1.25–1.39)
Breast cancer	–	0.76 (0.73–0.79)
Colorectal cancer	1.21 (1.16–1.27)	0.91 (0.86–0.96)
External causes	2.67 (2.58–2.78)	1.41 (1.35–1.48)

Mortality rate ratio = $\frac{\text{Lowest quartrile}}{\text{Highest quartrile}}$ of socioeconomical status

[a] 95% confidence interval

Conclusions

The United States now has nearly a half century of experience of improving air quality under the federal statute, the Clean Air Act, first enacted in 1963. The amendments of 1970, 1977 and 1990 substantially strengthened the CAA. Remarkable progress has been made in improving air quality as assessed using multiple criterion. The establishment of National Ambient Air Quality Standards for criteria pollutants by the EPA and the implementation programs of the individual States have contributed significantly to that success. Every decade from 1970 to the present has seen major actions with regard to the NAAQS and, in general, more stringent Standards. In many instances, Standards have been attained or nearly attained, and then a new more stringent Standard has been introduced. As some have said, we were almost there and then they moved the goal posts, i.e. lowered the Standards.

Now, more than at any time in the past, the policy judgment question must be asked "How low is low enough?" for each of the NAAQS. In my opinion, the guidance of Justice Breyer provides the Administrator broad latitude to make policy judgments consistent with our common goal of enhancing the health of all Americans.

Whatever path is chosen to go forward, there will remain a need for policy judgments informed by the best available scientific information. In creating new scientific information, I urge scientists to think broadly and adopt a strong comparative health benefit orientation. For example, when conducting epidemiological investigations, include multiple air pollutants and other factors, including socio-economic status that may influence the health endpoints being evaluated. Then report on all of the tested associations, not just the results for a single air pollutant. The resulting broader base of knowledge will allow Society to make decisions as to what actions will yield the most improvement in health at the lowest net cost to Society.

When future Integrated Science Assessment Documents are prepared, I urge that they include information that will help put the reported health effects of the specific pollutant in context. One approach to this might be the development of a generic document that reviews current knowledge on the multiple factors that influence morbidity and mortality from respiratory and cardiovascular disease, the major health outcomes for key criteria pollutants. This information could then be used in multiple Policy Assessment Documents. Both the Integrated Science Assessment and Policy Assessment Documents should more clearly identify and characterize the health effects role of the specific pollutant under consideration as well as the role of co-pollutants and other factors influencing the health outcomes evaluated. Policy Assessment Documents need to include "determinate criterion for drawing lines" as called for by the DC Circuit Court in its *American Trucking Associations v. EPA* (1999) opinion. These are needed to provide a clearer basis for the Administrator's policy judgments on the level and form of the Standard. These criteria, along with a strong comparative health context, should provide an improved basis for the Administrator's policy decisions.

I also strongly urge the CASAC to focus on the scientific rigor of the scientific content and analyses in the Integrated Science Assessment and Policy Assessment Document, and avoid the temptation of offering policy judgments as to a specific upper-bound level and form of the Standard or what they view as acceptable ranges. If CASAC cannot avoid this temptation to stay out of the "policy judgment thicket," then it needs to be clear as to the specific scientific knowledge that informs their personal policy preferences. CASAC is required to comment to the Administrator under CAA § 109(d)(2)(B) "on any new national ambient air quality Standards and revisions of existing criteria or Standards as may be appropriate." However, in offering comments, CASAC needs to very carefully articulate where CASAC scientific interpretations leave off and CASAC policy judgments begin. Moreover, it is important for EPA Administrators to recognize they need not be bound by CASAC's specific policy preferences or range of policy preference outcomes. While the CASAC members are citizens and are certainly entitled, just like any citizen, to have personal preferences as to policy outcomes, CASAC members, acting in that role, should not view themselves as broadly representative of Society at large.

It is critically important that EPA Administrators recognize, as Administrator William Ruckelshaus so clearly did in 1983, that Standards cannot be set solely on science and that the ultimate decision on a level and form of a Standard necessarily reflects policy judgments. Administrators should not seek to find "scientific cover" for these policy judgments in the deliberations offered by CASAC. If this is done, it has the potential to transform the Clean Air Scientific Advisory Committee into a de facto Clean Air Standards Setting Committee, thereby usurping the policy role of the Administrator. I do not think that is consistent with the language of the CAA. The Administrator, as a public official appointed by the President and confirmed by the Senate, is expected to have a broad perspective reflective of all of Society, not just a specific scientific constituency, when making policy judgments in setting National Ambient Air Quality Standards.

Declaration of Interest I have participated, beginning in the mid-1970s, as a member of numerous CASAC Panels providing advice to the EPA Administrator on the setting of the NAAQS for all the criteria pollutants. I served as Chair of CASAC in 1988–1992 when the debate began on shifting the averaging time for the ozone Standard from 1 to 8 h. I served on the CASAC PM Panels that provided advice on the $PM_{2.5}$ Standards promulgated in 1997 and 2006. I served on the CASAC Ozone Panel that provided advice on the Standard promulgated in 1997. I did not serve on the CASAC Ozone Panel that provided advice to the EPA Administrator on the Standard promulgated in 2008. However, I did follow that activity closely and offered comments to CASAC and EPA on the science informing the Administrator's judgments on the Ozone NAAQS. The views I share in this paper are my own professional views based on three decades of experience participating in the NAAQS setting process. I regularly serve as an advisor to both public and private organizations on air quality issues. This includes the American Petroleum Institute (API) and various companies in the energy and transportation sectors. The views I have expressed are not necessarily those of the API or any organization I advise.

References

American Trucking Associations, Inc. vs U.S. Environmental Protection Agency (1999) DC Circuit, May 19, 1999

Bachmann J (2007) Will the circle be unbroken: a history of the U.S. National Ambient Air Quality Standards. J Air Waste Manage Assoc 57:652–697

Berry MA (1984) A method for examining policy implementation: a study of decision-making for the national ambient air quality standards, 1964–1984; U.S. Environmental Protection Agency, Research Triangle Park, NC

Breyer S (1982) Regulation and reform. Harvard University, Cambridge, MA

Breyer S (1993) Breaking the vicious circle: toward effective risk regulation. (Based on 1992 Oliver Wendell Holmes lectures at Harvard Law School). Harvard University, Cambridge, MA

CAA (1963) The Clean Air Act of 1963

CAA (1970) The Clean Air Act amendments of 1970-P.L. 91–04 (December 31, 1970), Plus technical amendments made by P.L. 92–157 (November 18, 1971)

CAA (1977) The Clean Air Act amendments of 1977

CAA (1990) The Clean Air Act amendments to 1990, Public Law No. 101–549; 104 STAT 2399, 1990

Clinton WJ, President (1993) Executive Order 12866—regulatory planning and review. Fed Regis 58:51735–51744

Clinton WJ, President (1997) Memorandum for the administrator of the environmental protection agency implementation of revised air quality standards for ozone and particulate matter

Emanuel R (2009) Assistant to the President and Chief of Staff, Memorandum to the Heads of Executive Departments and Agencies. Fed Reg 74(13):4435–4436

Henderson RF (2005) Letter to administrator on particulate matter standard from CASAC Chair, Rogene Henderson to EPA Administrator Stephen Johnson, June 6, 2005

Henderson RF (2006a) Clean air scientific advisory committee recommendations concerning the proposed national ambient air quality standards for particulate matter: Letter from CASAC Chair, Rogene Henderson to EPA Administrator Stephen Johnson, March 21, 2006

Henderson RF (2006b) Clean air scientific advisory committee's review of the agency's draft ozone staff paper: letter from CASAC chair, Rogene Henderson to administrator Stephen Johnson dated October 24, 2006, EPA-CASAC-07-001

Henderson RF, Cowling EB, Crapo JD, Miller FJ, Poirot RL, Speizer FE, Zielinska B (2006c) Clean air scientific advisory committee recommendations concerning the proposed national ambient air quality standards for particulate matter: letter to EPA administrator Stephen L. Johnson, EPA-CASAC-LTR-06-002, March 21, 2006; EPA clean air scientific advisory committee, Washington, DC

Henderson RF (2007) Clean air scientific advisory committee's review of the agency's final ozone staff paper: letter from CASAC Chair, Rogene Henderson to administrator Stephen Johnson dated March 26, 2007, EPA-CASAC-07-002

Henderson RF (2008) Clean air scientific advisory committee recommendations concerning the final rule for the national ambient air quality standards for ozone: letter from CASAC Chair, Rogene Henderson, to EPA Administrator Stephen Johnson, April 7, 2008, EPA-CASAC-08-009

Kennedy D (2005) Risk versus risk. Science 309:2137, 30 September 2005

Library of Congress (1974) A legislative history of the clean air amendments of 1970, Vol. 1; Serial number 93–18: Environmental Policy Decision, Congressional Research Service, Library of Congress, Washington, DC

Martin K (2008) Building on the last ozone NAAQS Review: key policy relevant issues. Presentation to EPA Workshop on Science/Policy Issues for Next Periodic Review of Ozone NAAQS, October 29, 2008

Martineau R, Novello D (2004) The Clean Air Act handbook. 2nd edition

McClellan RO (1999) Human health risk assessment: a historical overview and alternative paths forward. Inhal Tox 11:477–518

McClellan RO (2007) Comments on national ambient air quality standards for ozone: proposed rule. Submitted on October 9, 2007 to Docket ID No. EPA-HQ-OAR-2005-0172

McClellan RO, Frampton MW, Koutrakis P, McDonnell WF, Moolgavkar S, North DW, Smith AE, Smith RL, Utell MJ (2009) Critical considerations in evaluating scientific evidence of health effects of ambient ozone: a conference report. Inhal Toxicol 21(2):1–36

McClellan RO (2010a) Comments on proposed "Reconsideration" national ambient air quality standard for ozone for presentation at the public meeting. U.S. Environmental Protection Agency, Arlington, VA, February 2, 2010

McClellan RO (2010b) Comments on the proposed "Reconsideration" national ambient air quality standard submitted on March 9, 2007 to the Ozone Docket

McClellan RO (2010c) Hazard and risk assessment and management. In: Jon Ayres G, Roy Harrison M, Maynard R, Gordon Nichols L (Eds.), Textbook of Environmental Medicine Chapter 4 1st Ed., Hodder Education: London, UK pp 59–88

McClellan RO (2011) Written statement for consideration by the clean air scientific advisory committee (CASAC) Ozone review panel at the panel's teleconference meeting on February 18, 2011

New York Times (1997) Clinton sharply tightens air pollution regulations despite concern over costs. NY Times, June 26, 1997

NRDC vs Train (1976) 411F. Supp. 864 (S.D.N.Y.) aff'd 545F.2d 320 (2nd Cir. 1976)

Peacock M (2006) Memorandum from Marcus Peacock, deputy administrator to Dr. George Gray and Bill Wehrum. Process for reviewing the national ambient air quality standards, December 7, 2006; available at http://www.epa.gov/ttn/naaqs/memo_process_for_reviewing_naaqs.pdf (accepted 2007)

Pedersen WF (1975) Formal records and informal rulemaking. Yale Law J 38:38–88

Rhomberg LR, Goodman JE, Haber LT, Dougson M, Andersen MT, Klaunig JE, Meek B, Price PS, McClellan CSM (2011) Linear low-dose extrapolation for non-cancer health effects is the exception, not the rule. Crit Rev Toxicol 41(1):1–19; available at http://informahealthcare.com/toc/txc/41/1

Samet JM (2011) Clean air scientific advisory committee (CASAC) Response to charge questions on the reconsideration of the 2008 Ozone national ambient air quality standard. EPA-CASAC-11-004. Letter from CASAC Chair, Jonathan M. Samet to Administrator Lisa P. Jackson, March 30, 2011

Steenland K, Hu S, Walker J (2004) All-cause and cause-specific mortality by socioeconomic status among employed persons in 27 U.S. States, 1984–1997. Am J Public Health 94:1037–1042

Taube G (1995) Epidemiology faces its limits. Science 269:164–169

U.S. Environmental Protection Agency (1971) National primary and secondary ambient air quality standards. Fed Reg 36:8186

U.S. Environmental Protection Agency (1977a) Air quality criteria for lead (released December 14, 1977). EPA/600/8-77/017

U.S. Environmental Protection Agency (1977) Lead ambient air quality standard proposal. Fed Regis 42:63076

U.S. Environmental Protection Agency (1978) National ambient air quality standard for lead. Fed Regis 43:46246

U.S. Environmental Protection Agency (1997a) National ambient air quality standard for particulate matter. Final Rule Fed Reg 62:38652–38760, July 18, 1997

U.S. Environmental Protection Agency (1997b) National ambient air quality standard for ozone. Final Rule Fed Reg 62(138):38855–38896, July 18, 1997

U.S. Environmental Protection Agency (2004) Air quality for particulate matter (October 2004). 600/P-99/002aF-bF, Washington, DC

U.S. Environmental Protection Agency (2005) Review of the national ambient air quality standards for particulate matter: policy assessment of scientific and technical information (December 2005). EPA-452/R-05-005a, Washington, DC

U.S. Environmental Protection Agency (2006a) National ambient air quality standards for particulate matter: proposed rule. Fed Regis 71:2620–2708

U.S. Environmental Protection Agency (2006b) National ambient air quality standards for particulate matter: final rule. Fed Regis 71:61144–61233

U.S. Environmental Protection Agency (2006c) Air quality criteria for ozone and related photochemical oxidants (Final). Washington, DC: 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

U.S. Environmental Protection Agency (2007b) Ozone health risk assessment for selected urban areas, EPA-452/R-07-009

U.S. Environmental Protection Agency (2007) National ambient air quality standards for ozone: proposed rule 40 CFR part 50. Fed Regis 72:37818–37919

U.S. Environmental Protection Agency (2007d) 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. Fed Regis 73:16436–16514

U.S. Environmental Protection Agency (2010a) National ambient air quality standards for ozone. Proposed Rule Fed Reg 75:2938, January 19, 2010

U.S. Environmental Protection Agency (2010b) Primary national ambient air quality standards for nitrogen dioxides: final rule. Fed Reg 75:6474, February 9, 2010

U.S. Environmental Protection Agency (2010c) Primary national ambient air quality standard for sulfur dioxide: final rule. Fed Regis 75:35520–35603, June 22, 2010

U.S. Environmental Protection Agency (2011a) Integrated science assessment for ozone and related photochemical oxidants (First External Review Draft). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-10/076A

U.S. Environmental Protection Agency (2011b) Notification of Public Teleconference of the Clean Air Scientific Advisory Committee (CASAC): Ozone Review Panel for Reconsideration of the 2008 National Ambient Air Quality Standard. Fed Reg 76:10895, February 28, 2011

Wegman LN (2011) Solicitation of CASAC Advice on EPA's Reconsideration of the 2008 Primary Ozone National Ambient Air Quality Standard. Letter to Holly Stallworth. Designated Federal Official, CASAC, January 25, 2011

White RH, Cote I, Zeise L, Fox M, Dominici F, Burke TA, White PD, Hattis DB, Samet JM (2009) State-of-the-Science Workshop Report: Issues and Approaches in Low Dose-Response Extrapolation for Environmental Health Risk Assessment. Available at: http://www.ehponline.org/members/2008/1150211502.pdf. Environmental Health Perspectives 117(2): doi:10.1289/ehp.11502

Whitman vs American Trucking Associations (2001) 531 U.S. 457, 121 S. Ct. 903149L. Ed. 2d1

Wolff GT (1996) Letter to Carol M. Browner, U.S. Environmental Protection Agency Administrator. closure on draft Office of Air Quality Planning and Standards (OAPQS Staff Paper) (Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Science and Technical Information)

EXHIBIT B

Risk Analysis, Vol. 36, No. 9, 2016 DOI: 10.1111/risa.12674

Invited Commentary

Providing Context for Ambient Particulate Matter and Estimates of Attributable Mortality

Roger O. McClellan*

Four papers on fine particulate matter ($PM_{2.5}$) by Anenberg *et al.*, Fann *et al.*, Shin *et al.*, and Smith contribute to a growing body of literature on estimated epidemiological associations between ambient $PM_{2.5}$ concentrations and increases in health responses relative to baseline notes. This article provides context for the four articles, including a historical review of provisions of the U.S. Clean Air Act as amended in 1970, requiring the setting of National Ambient Air Quality Standards (NAAQS) for criteria pollutants such as particulate matter (PM). The substantial improvements in both air quality for PM and population health as measured by decreased mortality rates are illustrated. The most recent revision of the NAAQS for $PM_{2.5}$ in 2013 by the Environmental Protection Agency distinguished between (1) uncertainties in characterizing $PM_{2.5}$ as having a causal association with various health endpoints, and as all-cause mortality, and (2) uncertainties in concentration—excess health response relationships at low ambient $PM_{2.5}$ concentrations below the majority of annual concentrations studied in the United States in the past. In future reviews, and potential revisions, of the NAAQS for $PM_{2.5}$, it will be even more important to distinguish between uncertainties in (1) characterizing the causal associations between ambient $PM_{2.5}$ concentrations and specific health outcomes, such as all-source mortality, irrespective of the concentrations, (2) characterizing the potency of major constituents of $PM_{2.5}$, and (3) uncertainties in the association between ambient $PM_{2.5}$ concentrations and specific health outcomes at various ambient $PM_{2.5}$ concentrations. The latter uncertainties are of special concern as ambient $PM_{2.5}$ concentrations and health morbidity and mortality rates approach background or baseline rates.

KEY WORDS: Clean Air Act; criteria pollutants; National Ambient Air Quality Standards; particulate matter; $PM_{2.5}$

The purpose of this commentary is to provide context and perspective for considering the contents and conclusions of four articles in this issue of *Risk Analysis* concerned with ambient fine particulate matter, 2.5 micron ($PM_{2.5}$) and estimates of $PM_{2.5}$ attributable mortality.

1. KEY ELEMENTS OF FOUR ARTICLES

Before offering my comments, I will briefly summarize what I view as key aspects of the four articles.

*Toxicology and Human Health Risk Analysis, Albuquerque, NM, USA; roger.o.mcclellan@att.net.

Anenberg *et al.*[1] provide a useful review of 12 air pollution health impact assessment tools that have been extensively used internationally. The tools use common data sources for the key inputs: (1) ambient $PM_{2.5}$ concentration–response association functions, (2) measured or estimated ambient concentrations of $PM_{2.5}$, (3) populations evaluated, and (4) baseline mortality rates. The models are all grounded in linear concentration–response functions.

Fann *et al.*[2] focus on the strengths and weaknesses of four research synthesis approaches to characterizing the long-term ambient $PM_{2.5}$ concentration–response functions. They note "whether

implicitly or explicitly, all require considerable judgment by scientists," an admonishment that should be heeded by both scientists and policymakers. Their focus is on linear models of ambient $PM_{2.5}$ concentration–response relationships. They provide some useful examples of estimated $PM_{2.5}$ attributable premature deaths based on different ambient concentration–response functions, estimates for which I will provide context.

Shin *et al.*[3] review meta-analysis methods for estimating the shape and uncertainty in the association between long-term exposure to ambient particulate matter (PM) and all-cause mortality. Their article considers both linear concentration–response models and alternative models extending to higher concentrations as required for some global applications. This is an especially important consideration when addressing the global range of $PM_{2.5}$ concentrations from current low ambient concentrations observed in countries such as the United States and Canada that have aggressively regulated air pollutants for half a century to countries like China and India with recent rapid industrialization, more limited regulations, and very high ambient concentrations of $PM_{2.5}$ and other air pollutants. Equally as important and not addressed by Shin *et al.* are the remarkable differences in various characteristics among the countries and within countries, including population characteristics such as baseline mortality rates, which are key inputs to the models.

The fourth article by Smith[4] illustrates the use of alternative approaches to calculating expected benefits of reducing the U.S. annual National Ambient Air Quality Standard for $PM_{2.5}$ from 15 to 12 $\mu g/m^3$. This article contains useful examples of the marked differences in estimates of avoided premature deaths dependent on the assumptions used in the calculations, including whether deaths are projected to occur below the current U.S. annual standard of 12 $\mu g/m^3$. I will provide context to those calculated estimates of avoided premature deaths.

2. THE CLEAN AIR ACT AS CONTEXT

The subject matter of the articles is grounded in the Clean Air Act (CAA) originally passed in 1963,[5] extensively amended in 1970[6] and again in 1990.[7] The CAA is the primary legislative basis for addressing air quality in the United States. Key sections of the CAA that require the U.S. Environmental Protection Agency (EPA) to set National Ambient Air Quality Standards (NAAQS) for certain air pollutants found across the United States and attributable to multiple sources, based on scientific criteria; hence, in common usage they are called criteria pollutants. The CAA identifies two types of NAAQS. Primary standards are intended to protect public health, including protection of "sensitive" populations such as asthmatics, children, and the elderly. The primary standards are the focus of this commentary. The CAA also calls for secondary standards to protect public welfare, which includes visibility and damage to animals, crops, vegetation, and buildings. Bachman[8] reviewed the long history of the NAAQS, a paper that should be read by all who are interested in this topic.

It is useful to recall that passage of the CAA was motivated by widespread recognition in the 1950s and 1960s that the United States had serious air quality problems arising from a marked increase in industrial activity during and after World War II. In addition, it was recognized that air quality was being increasingly impacted by expanded use of motor vehicles. It was generally accepted that poor air quality was impacting the health of the populace. Initial attempts to control air pollution were grounded in local and state legislation. It soon became apparent that these actions were inadequate; hence, the CAA, as passed in 1963, was national in scope. Indeed, it specified the creation of a National Air Pollution Control Agency. This agency would ultimately become the "air office" component of the U.S. EPA when it was created on December 2, 1970.

The CAA amendments of 1970 substantially elevated the federal role in improving air quality, including the setting of NAAQS. The amended CAA (1970) delegates to the EPA Administrator responsibility for policy decisions on setting the four elements of each NAAQS (the indicator such as $PM_{2.5}$, the averaging time such as annual or 24 hour, the concentration, and the statistical form used to determine when the standard is attained). It is important to recognize that the CAA gives the EPA Administrator broad policy-making discretion for setting each NAAQS. The primary or health-based NAAQS are standards set so as to provide requisite protection, neither more nor less stringent than is necessary to protect public health, with an adequate margin of safety. The CAA does not specify a quantitative goal for setting each NAAQS based on some specific level of health protection, i.e., an acceptable level of risk. Thus, the level of risk protection embedded in each NAAQS is a policy judgment delegated to the EPA Administrator. Further, the U.S. Supreme Court in

Whitman vs. American Trucking Association[9] ruled that in setting the NAAQS, the Administrator cannot consider the costs of achieving the standards.

The six original criteria pollutants were PM, photochemical oxidants, carbon monoxide, sulfur dioxides, nitrogen oxides, and hydrocarbons. It was later determined that the hydrocarbons were more appropriately addressed as individual pollutants under the hazardous air pollutants section of the CAA. Legal action in the 1970s initiated by the National Resources Defense Council forced EPA to list lead as a criteria air pollutant. NAAQS have been set for each of the criteria pollutants and the science undergirding each NAAQS periodically reviewed. Most reviews have concluded with revision of the NAAQS. In addition, a national network of monitors has been established, primarily for regulatory compliance purposes. These monitors also provide the data that have been key to the conduct of most long-term epidemiological studies.

PM is a generic term for a broad class of chemically and physically diverse substances that exist as discrete particles (liquid droplets or solids) over a range of sizes in the ambient air. It is important to recall that the original NAAQS for PM set in 1971 used "total suspended particles" (TSP) as an indicator. TSP samples are collected with a high volume sampler and include particles up to 25–45 microns in size. Standards were set for both 24-hour and annual averaging time. The latter was set at 75 $\mu g/m^3$, annual geometric mean form. In the discussion that follows, the focus will be on the annual standard. After an extensive review process initiated in the late 1970s, the PM NAAQS was revised in 1987 with the TSP indicator replaced with a particulate matter, 10 microns (PM_{10}) indicator. It is important to recognize that the PM_{10} fraction is included within the size range of TSP samples. The new annual PM_{10} NAAQS was set at 50 $\mu g/m^3$ and the form changed to an annual arithmetic mean, averaged over three years.

A contentious review concluded in 1997 resulted in a revision of the PM NAAQS with the addition of a $PM_{2.5}$ indicator despite there being very limited $PM_{2.5}$ ambient concentration–response data available for setting the NAAQS with 2.5 micron $PM_{2.5}$ indicators. Keep in mind that the $PM_{2.5}$ fraction is included within the size range of the PM_{10} fraction. The $PM_{2.5}$ annual NAAQS was set at 15 $\mu g/m^3$, annual arithmetic mean, averaged over three years. To give impetus to the adoption of a $PM_{2.5}$ indicator, one EPA official commented: "If you want monitoring data on $PM_{2.5}$ for epidemiological studies, you need to support setting a NAAQS for $PM_{2.5}$, we only monitor what is regulated." In 2006, after another contentious review, the PM NAAQS was revised with a reduction in the 24-hour standard from 65 to 35 $\mu g/m^3$ and no change in the annual standard. In 2012, after another review, the PM standard was again revised with a reduction in the primary annual NAAQS to 12 $\mu g/m^3$, annual arithmetic mean averaged over three years. The next cycle of review of the PM NAAQS is already underway. If the agency were to conform with a five-year review cycle, the next review should be concluded by 2018. The agency has already acknowledged that it will not meet that schedule and instead has announced a schedule for release of the final PM rate in 2021.

In my opinion, the changes in the annual PM NAAQS over the decades have been driven largely by (1) improved scientific knowledge on the role of particle size governing the deposition and retention of airborne particles, hence the serial shift from a TSP to PM_{10} to $PM_{2.5}$ indicator, and (2) improved knowledge from epidemiological studies of human populations such as those under discussion in the four articles. The policy decision of the EPA Administrator on the level and form of the NAAQS for PM has largely been informed by the information from epidemiological studies.

All of the PM NAAQS set to date are based on mass concentration and the assumption that all of the PMs in each size fraction are of equal toxicity on a mass basis. This assumption needs careful review in the current PM review cycle.

3. HISTORIC CHANGES IN $PM_{2.5}$ AND MORTALITY

To provide context for considering the contents of the four articles, it is useful to consider the substantial historic changes in ambient $PM_{2.5}$ and mortality rate in the United States. One of the major long-term studies of the association between ambient PM and mortality is the Harvard Six Cities Study, a study conceived by Professor Benjamin Ferris in the 1970s when revision of the NAAQS set in 1971 was under review. Updated findings from this study have been periodically published. The recent paper by Lepeule *et al.*[10] provides a useful summary of the changes in ambient $PM_{2.5}$ concentrations in the six cities from the mid 1970s through 2009. The range of ambient concentrations shown (Fig. 1) is a reasonable representation of the downward trend in

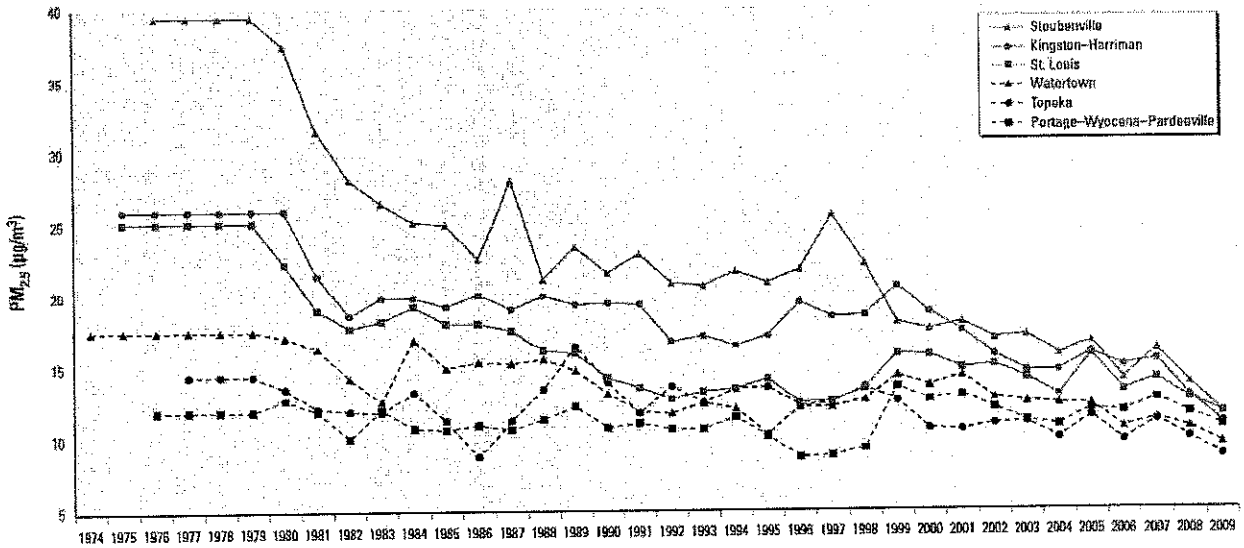


Fig. 1. Annual mean $PM_{2.5}$ levels during 1974–2009 in the Harvard Six Cities Study. (Adapted from Lepeule *et al.*[10] The data points pre-1997 for $PM_{2.5}$ have been extrapolated from TSP and PM_{10} measurements.)

urban areas seen across the United States over this time period. In reviewing the figure, keep in mind that the PM indicator from 1971 to 1978 was TSP and from 1978 to 1997 was PM_{10} with the $PM_{2.5}$ indicator added in 1997. The $PM_{2.5}$ concentrations shown in the figure for the earliest years are extrapolations from other indicators. The reductions in ambient $PM_{2.5}$ are impressive, especially for the three cities that originally had concentrations of 25 $\mu g/m^3$ and higher. It is reasonable to assume that these cities experienced even higher concentrations of $PM_{2.5}$ and coarse particles (PM_{10} minus $PM_{2.5}$) at earlier times.

During the last three-quarters of a century, there have also been impressive improvements in mortality rates across the United States, with continuous reductions in crude death rates and even more impressive reductions in age-adjusted death rates.[11] Data for the period 1960–2010 are shown in Fig. 2.[12] It is important to note that these are national statistics with important substantial differences in both crude and age-adjusted death rates (deaths per 100,000 population) among different states and racial groups. For example, the age-adjusted death rate (all causes) in 2010 ranged from 590 in Hawaii to 962 in Mississippi.

Further context is provided by the data in Table I as to cause of death for mortality in the United States in 2010.[12] Consideration of these multiple causes of death provides insight into potential opportunities

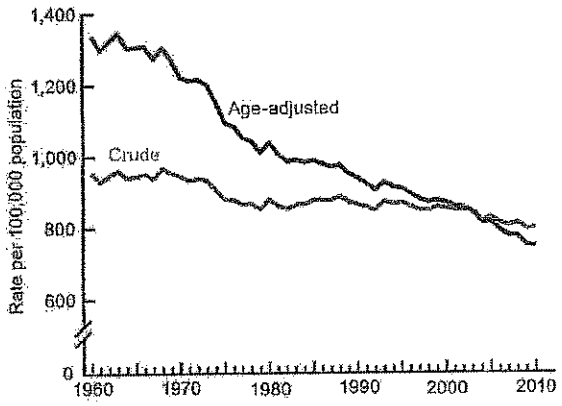


Fig. 2. Crude and age-adjusted death rates: United States, 1960–2010. (Adapted from Murphy *et al.*[12])

for improving the health of the U.S. population, our ultimate goal.

4. COMMENTER'S BACKGROUND FOR CONTEXT

It is important to recognize that provision of any context, to a large extent, is dependent on the commenters' backgrounds and how they view the

Table I. Causes of Death for the United States for 2010 by Major Causes[12]

Rank	Cause of Death (Based on ICD-10, 2004)	Number
...	All causes	2,468,435
1	Diseases of heart	597,689
2	Malignant neoplasms	574,743
3	Chronic lower respiratory diseases	138,080
4	Cerebrovascular diseases	129,476
5	Accidents (unintentional injuries)	120,859
6	Alzheimer's disease	83,494
7	Diabetes mellitus	69,071
8	Nephritis, nephrotic syndrome, and nephrosis	50,476
9	Influenza and pneumonia	50,097
10	Intentional self-harm (suicide)	38,364
11	Septicemia	34,812
12	Chronic liver disease and cirrhosis	31,903
13	Essential hypertension and hypertensive renal disease	26,634
14	Parkinson's disease	22,032
15	Pneumonitis due to solids and liquids	17,011
...	All other causes	483,694

application of the work being reviewed. The context and perspective I offer is grounded in my experience as a scientist, research manager, and advisor on the use of science to inform public policy decisions. I have been studying the health effects of airborne materials for over half a century, initially focusing on radioactive materials, as might be released in a nuclear reactor accident, and later on airborne emissions from various energy technologies, especially diesel compression ignition engines. Soon after passage of the CAA, I began advising both public agencies and private organizations on air quality issues at the interface between science and public policy. Much of that activity has involved the setting of NAAQS for criteria air pollutants, including PM and implementation of strategies to attain the NAAQS. This service included chairing the EPA's review committee for the first criteria document on airborne lead and later the EPA Clean Air Scientific Advisory Committee (CASAC) and service on the CASAC Panels that reviewed the science undergirding the 1987, 1997, and 2006 revisions of the PM NAAQS. I offered independent comments on the 2013 revision.

Based on my personal experience in the NAAQS setting process, I am firmly convinced that science should inform the policy decisions that are required in the setting of the NAAQS.[13] However, a corollary is that both scientists and policymakers should recognize that the science alone is not sufficient for making policy decisions. This is particularly the case in the absence of a quantitative goal or target for acceptable risk. The alternative approach embedded in the CAA is a policy judgment by the EPA Administrator as to how low is low enough. Tensions develop when scientists want to enter the policy arena and specify numerical standards that implicitly involve policy judgments. Tensions also arise when policymakers cast their policy judgments as being dictated by the science and abdicate their policy judgment role. I addressed those issues in the paper "Role of Science and Judgment in Setting National Ambient Air Quality Standards: How Low is Low Enough?"[13]

The passage of the CAA had substantial impact on the research enterprise in the United States, with substantial federal funding provided for investigation of pollutants from their movement from their sources at smoke stacks and tail pipes through the atmosphere to people and the development of an improved understanding of the health effects of airborne pollutants. A national network of monitors has been deployed, primarily for regulatory compliance purposes and secondarily for research purposes. Substantial investments of public and private funds have been made to develop and improve a wide range of technologies to reduce emissions of both regulated and nonregulated air pollutants from various sources. It is widely acknowledged today by multiple parties, the public, government agencies, industry, and politicians that the regulatory programs grounded in the CAA have had widespread positive impact. Air quality in the United States today is markedly improved from that observed in the 1970s and earlier. This leads to a critical question today as to what extent current air quality has any adverse impact on human health and, if so, are even more stringent NAAQS required? The first three articles under consideration address the science that informs policy decisions on the question posed. The fourth article by Smith[4] is at the interface of the science and policy. Some readers may be alarmed by my raising the issue of whether current air quality in the United States has adverse health impacts and requires more stringent standards. In my opinion, addressing that complex issue is at the interface of science and policy and is one reason why the four articles and related commentaries should be of interest to a wide audience of scientists, policymakers, and the public.

5. EVALUATING CAUSALITY

A critical issue related to assembling, integrating, synthesizing, and communicating the science on the health effects of $PM_{2.5}$ revolves around whether there is a "causal" link between exposure to ambient $PM_{2.5}$ and a range of health endpoints including all-cause mortality and specific causes of death such as ischemic heart disease, stroke, chronic obstructive pulmonary disease, and lung cancer. To aid in addressing this issue in an organized way, the EPA has developed a five-level hierarchy that classifies the overall weight of evidence drawn from integration of evidence across epidemiological, controlled human exposure studies, and toxicological studies and the related uncertainties that ultimately influence our understanding of the evidence. The five categories are: (1) causal relationship, (2) likely to be causal relationship, (3) suggestive of a causal relationship, (4) inadequate to infer a causal relationship, and (5) not likely to be causal relationship.[14] This approach is analogous to the hazard identification methodology widely used for decades in addressing cancer hazards of various agents. The *Federal Register* announcement of the National Ambient Air Quality Standards for Particulate Matter: Final Rule[15] has extensive discussion of the use of this qualitative categorical hazard hierarchy in informing the policy judgments supporting the decision (1) to lower the annual NAAQS for $PM_{2.5}$ from 15 to 12 $\mu g/m^3$ and (2) to retain the 24-hour averaging time NAAQS set at 35 $\mu g/m^3$ with a 98th percentile statistical form for attainment purposes.

It is noteworthy that this "causal" categorization process, by its very nature, emphasizes positive findings, which, in turn, emphasize the findings from studies at the highest ambient $PM_{2.5}$ concentrations. It is important to recognize that the categorization process does not rigorously address the equally important question of whether $PM_{2.5}$ at levels currently found in the United States have increased associated morbidity and mortality rates for specific health outcomes over and above baseline rates. That is a critical issue in the review of the science for a policy decision on any potential revision of the NAAQS for $PM_{2.5}$.

The issue of what ambient concentrations of $PM_{2.5}$ have a causal attributable effect on health outcomes such as an increase in all-cause mortality over and above background or baseline rates is not addressed by the five-level causal hazard hierarchy. This is a separate and extremely important issue. It is my opinion that many scientists, perhaps including some of the authors of the four articles, are confused and view the causal hazard hierarchy as extending to ambient $PM_{2.5}$ concentration–response functions.

Shin *et al.*[3] touch on this issue when they note the lowest concentration studied in the American Cancer Society (ACS) cohort was 5.8 $\mu g/m^3$, the 5th percentile was 8.8 $\mu g/m^3$, and the 95th percentile is below 20 $\mu g/m^3$. They note "reliable estimates of risk from the available studies can only be made using the data in the 5th to 95th percentile of exposure, i.e., estimates of the shape in the lower 5th and upper 95th percentile are both imprecise and likely to be inaccurate." I question the implication that the statistical association between ambient concentrations of $PM_{2.5}$ and excess risk is equally reliable over the full range from the 5th to the 95th percentile of $PM_{2.5}$ concentrations. It was disappointing that Shin *et al.*[3] did not more rigorously address the basis for their focus on the 5th percentile in view of EPA's approach to the last NAAQS revision.[15]

Specifically, it would have been of interest to readers if Shin *et al.*[3] had offered a rigorous critique of the related methodology used by the EPA Administrator to make the policy decision lowering the annual $PM_{2.5}$ NAAQS from 15 to 12 $\mu g/m^3$ effective from March 18, 2013.[15] In reaching that policy decision, the final rule stated: "In considering the evidence, the Policy Assessment recognized that NAAQS are standards set so as to provide requisite protection, neither more nor less stringent than necessary to protect public health, with an adequate margin of safety. This judgment ultimately made by the Administrator involves weighing the strength of the evidence and the inherent uncertainties and limitations of that evidence." As summarized in the Final Rule for the $PM_{2.5}$ NAAQS,[15] the Administrator gave special attention to four multicity studies for which distributional statistics of $PM_{2.5}$ ambient concentrations were available. This did not include the Harvard Six Cities Study, for which the Lepeule *et al.*[10] paper is the last update apparently, because the investigators would not release their data on ambient $PM_{2.5}$ concentrations for the populations studied in six cities. The Rule noted: "By considering this approach one could focus on the range of $PM_{2.5}$ concentrations below the long-term mean ambient concentrations over which we continue to have confidence in the associations observed in epidemiological studies (e.g., above the 25th percentile) where commensurate public health protection could be obtained for $PM_{2.5}$-related effects and, conversely, identify the range in the distribution below which our

confidence in the associations is appreciably less, to identify alternative annual standard levels." It is clear that this approach accepts the categorization of some long-term exposure studies as evidence of a causal or likely causal relationship for all-cause mortality; however, only above the 25 percentile of ambient $PM_{2.5}$ concentrations in the four studies. Most importantly, the EPA Administrator viewed the evidence below the 25th percentile as uncertain and not supportive of a causal or likely causal relationship. This contrasts with the conclusions of Shin *et al.*[3] It is very likely that this issue will be raised again in the next review of the $PM_{2.5}$ NAAQS. This is a critical issue at the interface between scientific information and policy choices. It is important to recognize that each review does not have to necessarily conclude with a revision of the NAAQS.

6. ASSOCIATIONS VERSUS CAUSALITY AT LOW $PM_{2.5}$ CONCENTRATIONS

All four of the articles most often referred to the "association" between ambient $PM_{2.5}$ and health responses. Unfortunately, the tone of three of the articles was that this association represented a causal relationship. As revealed in the earlier discussion of the EPA approach to setting the $PM_{2.5}$ (annual) NAAQS at 12 $\mu g/m^3$, it is important to not assume that causality extends to the lowest ambient $PM_{2.5}$ concentrations studied based on a linear model and the lowest ambient $PM_{2.5}$ concentrations studied. At a minimum, this issue deserves rigorous discussion and debate.

Unfortunately, none of the articles contain a robust discussion of the many biomedical uncertainties inherent in ambient $PM_{2.5}$ concentration–response associations over a range of ambient $PM_{2.5}$ concentrations. These uncertainties are multifold, including the official assumption in the last EPA review that all $PM_{2.5}$ is of equal toxicity on a mass basis. The assumption of equal toxicity is especially uncertain when one recognizes that PM reduction strategies have been highly effective in the United States over the past half-century in reducing mass emissions and reduced ambient concentrations of PM_{10} and $PM_{2.5}$. These reductions have resulted in a shift from PM resulting from direct emissions to PM formed from secondary reactions and associated changes in the chemical and size composition of PM. It is important to recognize that these changes are embedded in the ambient PM concentration data used in the major long-term epidemiological studies with the ambient PM for the earliest time periods in the studies being different from the ambient PM for the most recent updates of the studies. Unfortunately, speciated $PM_{2.5}$ data have rarely been obtained over long periods of time at multiple monitoring sites. Data on speciated $PM_{2.5}$ are necessary to test hypotheses on whether different $PM_{2.5}$ components have different potencies for causing an increase in different health effects. A closely related issue is whether ambient $PM_{2.5}$ concentration–response functions derived from the study of populations in one part of the United States are applicable to populations in other parts of the United States. The importance of this issue was underscored by the results reported by Zeger *et al.*[16] They found an association between increases in $PM_{2.5}$ and increases in mortality in the eastern and central regions of the United States and no evidence of an association in the western United States for the period 2000–2005. It is also important to recognize that the U.S. populations studied in recent decades were not likely exposed to PM of the composition and high concentrations encountered today in some countries such as China and India.

The Shin *et al.*[3] article has the most extensive discussion of the issue of causality. However, in my opinion, much of this discussion is quite simplistic and, indeed, naïve with regard to the actual complexity of disease processes. This is illustrated with the statement: "There is now experimental and clinical evidence that exposure to fine particulate matter causes biological responses such as oxidative stress leading to chronic inflammation, which in turn, can lead to increased mortality from chronic cardiovascular and respiratory disease and lung cancer, thereby shortening the lifespan." In my opinion, this is an excessively broad conclusion. I would agree that oxidative stress is one of the current fads in the biomedical sciences; however, such fads come and go. Unfortunately, disease processes are much more complex than this statement indicates, and a single step in complex multistep disease processes has rarely proved to be overwhelmingly dominant across a population afflicted with a particular disease. Shin *et al.*[3] use the term "causal models" at several places in their article, including reference to the paper of Pope *et al.*[17] These modeling exercises are useful; however, the models fall short of describing the myriad of complex steps by which responses over many decades to a single risk factor, such as $PM_{2.5}$, undefined as to chemical composition, cause a very small increase in the relative risk of death from a particular disease in a large population.

The Shin *et al.*[3] article contains what I view as an unjustified statement that: "There is no biological reason to believe that there exists a range in exposure for which no mortality risks exists (i.e., threshold)." It is noteworthy that the authors provided several figures in which data were plotted as hazard ratios or relative risk. The above quote apparently fails to recognize that the hazard ratio or relative risk of 1.0 is not an absence of mortality, it is the baseline mortality rate against which an increase in mortality attributable to the putative risk factor being examined is evaluated, in this case—$PM_{2.5}$—after attempting to control for all other risk factors potentially associated with the disease endpoint of concern. The diseases that are of concern for chronic exposure to $PM_{2.5}$ are very common causes of death (recall Table I) and arise from multiple risk factors. For deaths occurring late in life, many of these risk factors have interacted with normal biological processes, including damage, repair, and homeostatic processes, for decades throughout the individual's life. At the risk of sounding trite, life from conception to death is full of competing risks. The challenge for biomedical scientists, including statisticians, is to determine under what $PM_{2.5}$ exposure conditions over a lifetime of exposures there is a significant role for $PM_{2.5}$ in altering those complex processes and impacting morbidity and mortality rates. The challenge is even more difficult because many of the risk factors identified to date for the diseases of concern do have impact over the individual's total lifetime. As noted earlier (Fig. 1), there has been continuous improvement in mortality rates in the United States over the past half-century. Attempting to tease out the relative importance of a multitude of risk factors for this improvement in health is complex and beyond the scope of this commentary.

In this commentary, I have not discussed a growing body of evidence of a lack of influence of ambient $PM_{2.5}$ concentrations on mortality. An example is the paper by Greven *et al.*[18] that uses ambient $PM_{2.5}$ monitoring data for 2000–2006 and data on time of death and age for 18.2 million individuals in a Medicare cohort. They developed both national and local coefficients to examine trends. Based on the local coefficient alone, they were not able to demonstrate any change in life expectancy for a reduction in ambient $PM_{2.5}$. These results suggest the need for caution in using national values for estimating $PM_{2.5}$ attributable effects in specific regions of the United States, including California. In this regard, a number of studies have been developed on California populations, some of which suggest an absence of excess risk for recent ambient $PM_{2.5}$ concentration.[16,19]

It is well recognized by scientists and clinicians knowledgeable of the biology and pathobiology of the health endpoints of concern that none of the individual cases carry "markers" or any characteristics that allow $PM_{2.5}$ attributable cases to be distinguished from cases that are attributable to a myriad of other causes. The attribution of deaths associated with $PM_{2.5}$ exposure is done on a statistical and population basis. The statistical models used typically are proportional hazard models that estimate for the population a given portion of the cases over and above the baseline mortality rates attributed to other causes. To provide a context for considering the estimated $PM_{2.5}$ attributable deaths, it is always helpful to present the baseline mortality rate, which, as discussed earlier, varies with time, place, and population as influenced by multiple factors. I will return to that point later. In addition to showing the excess risk attributed to $PM_{2.5}$, it would be informative if the analysts also showed the excess risks estimated for other well-recognized risk factors, such as smoking and socioeconomic status, that must be controlled for in the analyses to develop reliable estimates of excess $PM_{2.5}$ attributable risks. This information would be valuable to the analysts and to other parties to help understand if the calculated results for $PM_{2.5}$ make sense. An array of attributable risk results for different risk factors also provides valuable context for policymakers and the public concerned with how best to positively impact human health. In my opinion, it is important to periodically recall the goal to improve public health; the regulation and control of specific risk factors such as $PM_{2.5}$ is just one means to that end.

7. EXPANDED PRESENTATION OF RESULTS TO PROVIDE CONTEXT

In this section, I will illustrate how an expanded presentation of results can provide useful context and perspective. Fann *et al.*[2] use their Fig. 3 to graphically illustrate the estimated "premature deaths avoided" based on different ambient $PM_{2.5}$ concentration–response functions. The focus is on comparison of the results using functions from the Harvard Six Cities Study[10] and the ACS study.[20] The graph also showed estimates developed from

Table II. Comparison of 2014 Estimated Premature Deaths Avoided Using Alternative Ambient $PM_{2.5}$ Concentration–Response Functions (Adapted from Fann *et al.*[2] and Fann [personal communication])

Source of Function	Baseline Mortality (Deaths)	Estimated Premature Deaths Avoided (Deaths 95% Confidence Interval)
Harvard Six Cities		
Lepeule *et al.*[10]	2,565,169[a]	10,373 (6,010, 14,698)
ACS		
Krewski *et al.* (2009)	2,565,169[a]	4,582 (3,334, 5,821)
Pooled experts	2,565169[a]	8,327 (1,492, 18,289)
Meta-analysis (beyond 2006)	2,565,169[a]	5,852 (2,527, 9,150)
Meta-analysis (through 2006)	2,565,169[a]	5,530 (3,287, 7,756)
Integrated exposure response	364,408[a]	3,931 (1,935, 4,241)

[a]All cause.
[b]Ischemic heart disease.

functions elicited from 12 experts, a meta-analyses of literature through 2006 and beyond 2006, and a pooling of the 12 experts based on all-cause mortality. Also shown is an estimate from an integrated exposure–response analysis for ischemic heart disease. In Table II, the original estimates of Fann *et al.*[2] are shown complemented by baseline mortality data added to provide context. In my opinion, showing the baseline mortality values helps the reader to understand this mathematical exercise. The table would be even more informative if it included the total population for 2014.

Smith[4] provides an excellent example of how the assumptions used in estimating benefits can have major impact on the results. In her paper, Table I showed the total risk reduction estimate (avoided premature deaths in 2020) for two approaches. One approach was the traditional approach used by EPA in developing regulatory impact analyses (RIAs). That approach assumes deaths are avoided irrespective of the ambient concentrations of $PM_{2.5}$. Table III yields 456 avoided deaths with the national concentration–response function that was developed by Krewski *et al.*[20] using the ACS cohort and 1,034 avoided deaths using the concentration–response function that was developed by Lepeule *et al.*[10] from the Six Cities Study. Smith[4] also gave lower estimates based on the approach that EPA used in the latest revision of the NAAQS for $PM_{2.5}$ described earlier in this commentary. As shown in Table III, the number of residual avoidable deaths is reduced to 21–47, dependent on the concentration–response function used and involves an impacted population of less than 1 million. Alleged benefits in the RIA, of 456–1,034 (or 460–1,000 using the RIA's rounding convention) avoidable deaths, disappear if one uses the qualitative policy judgment used by the EPA Administrator in revising the NAAQS for $PM_{2.5}$. Indeed, a strong argument can be made that there are no avoided $PM_{2.5}$ attributable deaths in California based on the report of Zeger *et al.*[16] Recall that they reported no finding of evidence of an association between ambient $PM_{2.5}$ and mortality in the western United States. They noted "this lack of association is largely because the Los Angeles Basin counties (California) have higher PM levels than other West Coast urban centers but not higher adjusted mortality rates." As an aside, California was the only state for which benefits of

Table III. Estimates of Avoided Premature Deaths in California in 2020 Estimated for $PM_{2.5}$ NAAQS with a Reduction in the Annual Standard from 15 to 12 $\mu g/m^3$ Projected Using BenMAP[21] and Smith (personal communication)

	Population		Baseline Mortality (#)		Avoided Deaths (#)	
	Krewski[a]	Lepeule[b]	Krewski	Lepeule	Krewski	Lepeule
	(30–99)	(25–99)	(30–99)	(25–99)	(30–99)	(25–99)
Not attaining/above margin (>13 $\mu g/m^3$)	763,104	875,086	7,574	7,681	21	47
Not attaining/in margin (>12–13 $\mu g/m^3$)	3,841,464	4,419,703	41,853	42,342	117	266
Already attaining (≤12 $\mu g/m^3$)	7,560,163	8,537,984	86,913	87,735	318	721
Total	12,164,732	13,832,773	136,340	137,758	456	1,034

[a]Krewski *et al.*[20] evaluate the population from age 30 to 99 years.
[b]Lepeule *et al.*[10] evaluate the population from age 25 to 99 years.

avoided mortality were projected to occur with a lowering of the $PM_{2.5}$ NAAQS. Other areas had already attained the $PM_{2.5}$ NAAQS. Again, the inclusion of the baseline population and mortality data helps provide context and perspective to the calculated benefits. Note that the population and baseline mortality values are based on actual data rather than hypothesized relationships and, thus, are much more certain than the calculated benefits. This broader array of data not only gives perspective to the calculated benefits, i.e., avoidable deaths, for a $PM_{2.5}$ standard, but invites questions as to where society at large can gain the greatest benefits in improved health.

To give a broader perspective to the estimated avoidable deaths, it is useful to recall Table I, which provides detailed mortality data by causes for 2010. As discussed earlier, consideration of calculated estimates of $PM_{2.5}$ attributable deaths along with an array of mortality data by multiple causes opens the door to a broader discussion of options for improving the health and quality of life for society at large moving beyond a singular focus on $PM_{2.5}$.

The above discussion has focused on providing information on three key inputs: (1) the population under consideration, (2) baseline mortality rate, and (3) the ambient $PM_{2.5}$ concentration–response functions (and the associated uncertainties at various $PM_{2.5}$ levels). It is also useful to have a more complete exposition of the ambient $PM_{2.5}$ data being used as input as illustrated by Smith.[4]

The above discussion also carries with it important implications for setting priorities for research that will help improve human health. Let me first address the adequacy of current models of ambient $PM_{2.5}$ concentration–response functions. In my view, the models currently available provide reasonable upper-bound estimates of $PM_{2.5}$ attributable mortality, i.e., more likely to overestimate than underestimate the true $PM_{2.5}$ attributable mortality. The estimated ambient $PM_{2.5}$ concentration–response functions and $PM_{2.5}$ attributable mortality calculated for those studies are likely related to the exposure of the populations over a lifetime beginning early in life, i.e., in the 1970s and earlier for the vast majority of the deaths. Ambient concentrations of $PM_{2.5}$ have steadily declined across the United States from that time to the present; recall Fig. 1. In addition, the U.S. age-adjusted death rate has steadily decreased, as shown in Fig. 2, related to many factors.

Let me quickly note that some individuals may suggest that improved air quality had a role in the observed reduced death rates. That may be true; however, I suggest the impact of $PM_{2.5}$ reductions is likely very small and difficult to tease out from the myriad of other factors that were likely involved in reducing mortality rates. To provide further perspective, it is useful to note the substantial impact of socioeconomic status on mortality[22] (Table IV). The mortality rate ratio for the lowest quartile over the highest quartile of socioeconomic status is high compared to small changes attributed to $PM_{2.5}$. It is obvious that many individual risk factors are included within socioeconomic states. All of these factors create "noise" that makes it challenging to identify any small signal attributed to $PM_{2.5}$. This speaks for caution in interpreting and using the small signals attributed to $PM_{2.5}$ in these statistical exercises.

Table IV. The Impact of Socioeconomic Status on Mortality Rate Ratio (Adapted from Steenland *et al.*[22])

Mortality	Men	Women
All causes	2.02 (1.95–209)[a]	1.29 (1.25–1.32)
Heart disease	1.88 (1.83–193)	1.84 (1.76–1.93)
Stroke	2.25 (2.14–2.37	1.53 (1.44–162)
Diabetes	2.19 (2.07–2.32)	1.85 (1.72–2.00)
COPD	3.59 (3.35–3.83)	2.09 (1.91–230)
Lung cancer	2.15 (2.07–2.23)	1.31 (1.25–1.39)
Breast cancer	–	0.76 (0.73–0.79)
Colorectal cancer	1.21 (1.16–1.27)	0.91 (0.86–0.96)
External causes	2.67 (2.58–2.78)	1.41 (1.35–1.48)

Note: Mortality rate ratio = mortality for lowest quartile of socioeconomic status. Mortality for highest quartile of socioeconomic status.
[a]95% Confidence interval.

The overall point I wish to make is that disease processes are very complex and are influenced by multiple risk factors. For any attempt to tease out the effects of a single risk factor, like $PM_{2.5}$, to be successful it needs to take account of the other risk factors. I urge the investigators who have focused their energy on $PM_{2.5}$ issues to broaden the scope of their research to give greater attention to identifying and characterizing multiple risk factors. In my opinion, this broader perspective offers the best opportunity for having a positive impact on the health of populations.

8. CONCLUSIONS

The information presented in the four articles and discussed here also has important implications for setting future PM NAAQS and for research to better understand mechanisms of disease causation, approaches to mitigation of disease, and treatment of

disease. A review of the data presented here, with a focus on the United States, indicates that any health effects attributable to $PM_{2.5}$ are quite small when considered in the context of the total disease burden. A corollary is the need for caution in advocating for more $PM_{2.5}$ focused research. In my opinion, a better return on societal investment is likely to come from a broader consideration of the complex pathways of disease causation common to multiple risk factors and, perhaps, amplified by certain risk factors.

REFERENCES

1. Anenberg SC, Belova A, Brandt J, Fann N, Greco S, Guttikunda S, Heroux ME, Hurley F, Krzyzanowski M, Medina S, Miller B, Pandey K, Ross J, Va Diagenen R. Survey of air pollution health risk assessment tools. Risk Analysis, 2016; 36(9):1718–1736.
2. Fann N, Gilmore EA, Walker K. Characterizing the long-term PM concentration-response function: Comparing the strengths and weaknesses of research synthesis approaches. Risk Analysis, 2015; 36(9):1693–1707.
3. Shin HH, Cohen AJ, Pope CA Ezzati M, Lim SS, Hubbell BJ, Burnett RT. Meta-analysis methods to estimate the shape and uncertainty in the association between long-term exposure to ambient fine particulate matter and cause-specific mortality over the global concentration range. Risk Analysis, 2016; 36(8) (this issue).
4. Smith AE. Inconsistencies in risk analyses for ambient air pollutant regulations. Risk Analysis, 2015; 36(9):1737–1744.
5. CAA. The Clean Air Act of 1963. 77 Stat, 392, 1963.
6. CAA. The Clean Air Act Amendments of 1970. 84 Stat. 1676, P.L. 91-604 (December 31, 1970). Plus technical amendments made by P.L. 92-157 (November 18, 1971), 1970.
7. CAA. Clean Air Act Amendments of 1990. 104 Stat. 2468, PL 101-549, 1990.
8. Bachman J. Will the circle be unbroken: A history of the U.S. National Ambient Air Quality Standards. Journal of the Air & Waste Management Association, 2007; 57:652–697.
9. Whitman vs American Trucking Associations. 531 U.S. 457, 121 Supreme Court 903149 L. Ed. 2d1, 2001.
10. Lepeule J, Laden F, Dockery D, Schwartz J. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard Six Cities Study from 1974 to 2009. Environmental Health Perspectives, 2012; 32(1):81–95.
11. Hoyert DL. 75 Years of Mortality in the United States, 1935–2010. National Center for Health Statistics Data Brief No. 88, Hyattsville, MD, March 2012.
12. Murphy SL, Xu J, Kochanek KO. Deaths: Final data for 2010. National Vital Statistics Report, 2013; 61(4):1–117.
13. McClellan RO. Role of science and judgment in setting National Ambient Air Quality Standards: How low is low enough? Air Quality, Atmosphere and Health, 2012; 5:243–258.
14. EPA (U.S. Environmental Protection Agency). Integrated Science Assessment for Particulate Matter: Final Report. Research Triangle Park, NC: National Center for Environmental Assessment – RTP Division, Office of Research and Development, EPA/600/R-08/139F. December 2009. Available at: http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_2007_isa_html, Accessed September 6, 2016.
15. EPA (U.S. Environmental Protection Agency). National Ambient Air Quality Standards for particulate matter: Final rule. Federal Register, 2012; 78:3086–3287.
16. Zeger SL, Dominici F, McDermott A, Samet JM. Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000–2005). Environmental Health Perspectives, 2008; 116:1514–1619.
17. Pope CA III, Burnett RT, Krewski D, Jerrett M, Shi Y, Calle EE, Thun MJ. Cardiovascular mortality and exposure to fine particulate matter from air pollution and cigarette smoke. Shape of the exposure–response relationship. Circulation, 2009; 120:941–948.
18. Greven S, Dominici F, Zeger S. An approach to the estimation of chronic air pollution effects using spatio-temporal information. Journal of the American Statistical Association, 2011; 106:396–406, 494.
19. Enstrom JE. Fine particulate air pollution and total mortality among elderly Californians 1983–2002. Inhalation Toxicology, 2005; 17:803–816.
20. Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA 3rd, Thurston G, Calle EE, Thun MJ, Beckerman B, DeLuca P, Finkelstein N, Ito K, Moore DK, Newbold KB, Ramsay T, Ross Z, Shin H, Tempalski B. Extended follow-up and spatial analysis of the American Cancer Society Study linking particulate air pollution and mortality. Research Report Health Effects Institute, 2009; 140:5–114, Accessed September 6, 2016.
21. BenMAP Version 4.0.67. Available at: http://www.epa.gov/air/benmap/download.html.
22. Steenland K, Hu S, Walker J. All-cause and cause-specific mortality by socioeconomic status among employed persons in 27 US states, 1984–1997. American Journal of Public Health, 2004; 94(6):1037–1042.

EXHIBIT C

Inhalation Toxicology, 17:803–816, 2005

ISSN: 0895-8378 print / 1091-7691 online
DOI: 10.1080/08958370500240413

Taylor & Francis
Taylor & Francis Group

Fine Particulate Air Pollution and Total Mortality Among Elderly Californians, 1973–2002

James E. Enstrom
Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California, USA, and Scientific Integrity Institute, Los Angeles, California, USA

Fine particulate air pollution has been associated with increases in long-term mortality in selected cohort studies, and this association has been influential in the establishment of air quality regulations for fine particles ($PM_{2.5}$). However, this epidemiologic evidence has been questioned because of methodological issues, conflicting findings, and lack of an accepted causal mechanism. To further evaluate this association, the long-term relation between fine particulate air pollution and total mortality was examined in a cohort of 49,975 elderly Californians, with a mean age of 65 yr as of 1973. These subjects, who resided in 25 California counties, were enrolled in 1959, recontacted in 1972, and followed from 1973 through 2002; 39,846 deaths were identified. Proportional hazards regression models were used to determine their relative risk of death (RR) and 95% confidence interval (CI) during 1973–2002 by county of residence. The models adjusted for age, sex, cigarette smoking, race, education, marital status, body mass index, occupational exposure, exercise, and a dietary factor. For the 35,789 subjects residing in 11 of these counties, county-wide exposure to fine particles was estimated from outdoor ambient concentrations measured during 1979–1983 and RRs were calculated as a function of these $PM_{2.5}$ levels (mean of 23.4 $\mu g/m^3$). For the initial period, 1973–1982, a small positive risk was found: RR was 1.04 (1.01–1.07) for a 10-$\mu g/m^3$ increase in $PM_{2.5}$. For the subsequent period, 1983–2002, this risk was no longer present: RR was 1.00 (0.98–1.02). For the entire follow-up period, RR was 1.01 (0.99–1.03). The RRs varied somewhat among major subgroups defined by sex, age, education level, smoking status, and health status. None of the subgroups that had significantly elevated RRs during 1973–1982 had significantly elevated RRs during 1983–2002. The RRs showed no substantial variation by county of residence during any of the three follow-up periods. Subjects in the two counties with the highest $PM_{2.5}$ levels (mean of 36.1 $\mu g/m^3$) had no greater risk of death than those in the two counties with the lowest $PM_{2.5}$ levels (mean of 13.1 $\mu g/m^3$). These epidemiologic results do not support a current relationship between fine particulate pollution and total mortality in elderly Californians, but they do not rule out a small effect, particularly before 1983.

Received 28 February 2005; accepted 21 June 2005.
The extended mortality follow-up and analyses presented in this article have been funded by the Electric Power Research Institute (EPRI). The entire funding history of CA CPS I prior to this analysis has been described elsewhere (Enstrom & Heath, 1999; Enstrom & Kabat, 2003). The author is responsible for all aspects of the article and declares no competing interests relevant to its contents. In the spirit of the Data Quality Act (OMB, 2003; Steinbrook, 2004), the author is willing to facilitate independent analysis of all the data used in the article. The author thanks Dr. Frederick W. Lipfert for proposing air pollution analyses with CA CPS I data and for contributing extensively to numerous versions of the text and tables, Dr. Ronald E. Wyzga for critiques of the article and for suggestions about making it as comprehensive and objective as possible, and Dr. Lingqi Tang for statistical assistance.
Address correspondence to James E. Enstrom, University of California, Jonsson Comprehensive Cancer Center, Los Angeles, CA 90095, USA. E-mail: jenstrom@ucla.edu

Many observational epidemiological studies have reported associations between air pollution from combustion sources and human health (Lipfert, 1994). During past severe air pollution events, such as the 1952 London fog incident (Logan & Glasg, 1953), extremely high concentrations of particulate air pollution were accompanied by major increases in coincident mortality. In more recent years, health effects have also been associated with much lower concentrations of particulate air pollution (Pope & Dockery, 1999). While much of the recent research has focused on short-term exposures, several studies suggest that long-term exposures may be more important. In particular, results from two major cohorts (Dockery et al., 1993; Pope et al., 1995, 2002) have shown significant mortality associations with outdoor concentrations of fine particles ($PM_{2.5}$, median aerodynamic diameter

less than 2.5 μm). Other cohort studies have also examined mortality associations with $PM_{2.5}$ and other pollutants (McDonnell et al., 2000; Lipfert et al., 2000, 2003), with somewhat different findings.

The major cohort studies have been used to support new national ambient air quality standards for fine particles issued by the U.S. Environmental Protection Agency (U.S. EPA, 1997). These standards are specific with respect to particle size, but not with respect to chemical composition. $PM_{2.5}$ is a variable mixture, rather than a defined chemical compound as in the case of gaseous air pollutants. Fine particles are generated mainly by combustion processes and their atmospheric sequelae, and all such particles measured by the approved methods are considered equally harmful. However, the chemical composition of airborne particulate matter varies appreciably across the nation and within metropolitan areas. Although national ambient air quality standards are intended to apply throughout the nation, it is not clear that the selected epidemiological studies on which those standards are based are equally applicable nationwide.

The associations of particulate air pollution with long-term mortality remain controversial (Phalen, 2002; Moolgavkar, 2005; Kaiser, 2005). This is in large part because the epidemiologic studies that have examined these health effects are subject to a number of methodological limitations (Greenbaum et al., 2001; Moolgavkar, 1996; Gamble, 1998; Krewski et al., 2000; Lipfert, 2003). Actual exposures to air pollution are difficult to determine accurately in large cohorts. Indeed, the exposure of each individual has not been directly measured in these studies, but has been assumed to equal the ambient outdoor $PM_{2.5}$ concentration for the individual's county or metropolitan area of residence. Also, one national cohort study has found largely negative associations between $PM_{2.5}$ and mortality (Lipfert et al., 2000, 2003).

California is a large, diverse state that has long been concerned about the health effects of air pollution and that has recently issued new stricter ambient $PM_{2.5}$ standards (CARB, 2003), based in large part on the national standards. However, no previous cohort study has focused on mortality with respect to measured $PM_{2.5}$ levels in California. This article used a large cohort of elderly Californians to examine in detail the relationship between $PM_{2.5}$ levels measured during 1979–1983 and mortality from all causes during 1973–2002.

METHODS

California Cancer Prevention Study

The California Cancer Prevention Study (CA CPS I) is the extended follow-up of the 118,094 California subjects from the original Cancer Prevention Study (CPS I) of 1,078,894 adults from 25 states. CPS I was initiated by the American Cancer Society (ACS) beginning in 1959, and CA CPS I has been independently conducted at the University of California, Los Angeles (UCLA), since 1991, as described in detail elsewhere (Enstrom & Heath, 1999; Enstrom & Kabat, 2003). The conduct of CA CPS I has been approved by the UCLA, Office for Protection of Research Subjects during this entire period. The subjects in this prospective cohort study were enrolled from October 1959 through February 1960 using a detailed four-page questionnaire. Surviving cohort members completed short questionnaires in late 1961, 1963, 1965, and 1972, and a two-page questionnaire in mid 1999. Deaths through 1972 were identified primarily by surviving study subjects and were confirmed with death certificates. The later deaths were identified primarily from computerized and manual matches with the California death file and the nationwide Social Security Death Index, using name and other identifying variables (Enstrom & Heath, 1999; Enstrom & Kabat, 2003). About 86% of the later deaths were identified on the California death file and the remainder were identified on the Social Security Death Index file. The only prior analysis of the CPS I cohort with respect to air pollution found no relationship between suspended particulate matter and lung cancer mortality during the 1960s and had no $PM_{2.5}$ data (Hammond, 1972; Hammond & Garfinkel, 1980).

This article analyzes those CA CPS I subjects who reported their cigarette smoking status in both the 1959 and 1972 questionnaires and who were alive as of January 1, 1973. Respondents to both questionnaires were traced more easily than those who responded only to the 1959 questionnaire. The 1972 questionnaire updated their cigarette smoking status, the most important confounding variable. The early years of follow-up (January 1, 1960–December 31, 1972) have not been included in this article because there are no statewide $PM_{2.5}$ data before 1979. Results for this early period and for the entire follow-up period since 1960 will be presented in a subsequent article dealing with other air pollutants. This analysis is limited to the 25 counties with the largest number of CA CPS I subjects, which ranged from 325 to 17,340 per county. About 95% of the CA CPS I subjects resided in these 25 counties. There were 49,975 traceable subjects alive as of January 1, 1973, of whom 39,846 died as of December 31, 2002. There were 35,789 traceable subjects alive as of January 1, 1983 in the 11 counties with $PM_{2.5}$ data, of whom 28,441 died as of December 31, 2002. An additional 2,735 subjects in these counties lost since January 1, 1973 (7.64%), have been omitted from further analysis.

The 1999 addresses for most of the traceable subjects alive as of January 1, 1999, were determined from a match with California driver's license (DL) identifying information, and about 33% of the subjects responded to a two-page smoking and lifestyle questionnaire that was mailed in mid 1999 to their DL address (Enstrom & Heath, 1999; Enstrom & Kabat, 2003). Based on the questionnaire information in late 1972, the 1999 DL address information, and the death information, the county of residence and county of death were determined for most subjects as of late 1972 and early 1999. The residential mobility of subjects was assessed by calculating the percentage of subjects who lived or died in the same county from 1972 to 1999.

1979–1983 $PM_{2.5}$ Data and 1973–2002 Mortality Data

The independent variable in this analysis is $PM_{2.5}$, as measured during 1979–1983 in 11 California counties by the EPA as part of the Inhalable Particulate Network (IPN) (Hinton et al., 1984, 1986), also known as the Inhalable Particle Monitoring Network (IPMN) (Sune, 1999; Pope et al., 2002). These data have been used in several previous epidemiological studies (Ozkaynak et al., 1987; Lipfert et al., 1988, 2000, 2003; Pope et al., 1995, 2002). In this article, the $PM_{2.5}$ data for each county were averaged over time and across the available monitoring stations, and are assumed to indicate the average long-term exposure level for all subjects in the county. No routinely measured $PM_{2.5}$ data in California exist before 1979 or during 1984–1998; routine statewide measurements in California were resumed in 1999. The average county-level $PM_{2.5}$ value was assigned to the traceable subjects alive as of January 1, 1973, based on their county of residence as of late 1972. This analysis was based on the deaths from January 1, 1973, to December 31, 2002, a 30-year follow-up period that includes the 5-yr period of the 1979–1983 $PM_{2.5}$ data. Additional analyses have been done for deaths from January 1, 1973, to December 31, 1982, and from January 1, 1983, to December 31, 2002. This latter period is roughly the same as the period (September 1, 1982–December 31, 1998) used in the recent national cohort study (Pope et al., 2002).

Analysis by Proportional Hazards Regression

The age- and sex-adjusted relative risk of death (RR) and 95% confidence interval (CI) were calculated using Cox proportional hazards regression, specifically the SAS PHREG procedure (SAS, 2004), including age at baseline in 1-yr intervals and sex, as a function of $PM_{2.5}$ level in units of 10 $\mu g/m^3$. This type of analysis is similar to the one used recently (Pope et al., 2002). Fully adjusted relative risks were calculated using a Cox model that includes age, sex, and eight potential confounding variables at baseline: cigarette smoking status (never, former as of 1959 and 1972, 1–9, 10–19, 20, 21–39, 40+ cigarettes per day as of 1972), race (white, nonwhite), education level (<12, 12, >12 yr), marital status (married, widowed, single, separated, divorced), body mass index (<20, 20–22.99, 23–25.99,

TABLE 1
Demographic and lifestyle characteristics in 1959 for California CPS I male subjects as of 1/1/1973 who resided in the 11 counties having 1979–1983 $PM_{2.5}$ measurements and who provided 10/1/1972 cigarette smoking status

Characteristic	1959 value (11 $PM_{2.5}$ counties)	1959 value (2 highest $PM_{2.5}$ counties)	1959 value (2 lowest $PM_{2.5}$ counties)	1959 value for 1999 respondents	1999 value for 1999 respondents
Mean level of 1979–1983 $PM_{2.5}$ ($\mu g/m^3$)	23.4	36.1	13.1		
Number of subjects alive as of 1/1/1973	16,296	1043	1040		
Lost to follow-up since 1/1/1973 (%)	4.432	2.987	4.840		
Number of subjects alive 1/1/1973 and not lost since 1/1/1973	15,574	1012	990	1314 (alive 1999)	1314 (alive 1999)
Age as of 1/1/1973 (mean, years)	65.7	67.1	64.5	58.4	58.4
Age as of 1/1/1983 (mean, years)	73.8	74.9	72.4	68.4	68.4
Race (% white)	98.4	99.0	97.5	98.5	98.0
Marital status (% married)	97.3	97.4	98.0	96.3	75.6
Education (% ≥12 yr)	71.8	70.7	79.8	90.3	92.6
Height (mean, inches)	69.4	69.5	69.9	69.9	69.3
Weight (mean, pounds)	173.0	172.7	174.5	172.9	168.9
History of serious diseases (% yes)	9.7	12.7	7.5	4.6	
Cancer	4.6	6.9	3.8	3.1	42.5
Heart disease	4.6	4.6	3.5	1.2	
Stroke	0.6	1.2	0.2	0.3	
Sick at the present time (% yes)	6.8	6.8	5.4	6.2	25.9
Occupation (% professional)	10.5	11.9	9.8	17.5	
Residence location (% urban)	98.1	99.2	98.6	98.1	
Exercise (% moderate or heavy)	72.5	73.9	76.1	67.1	61.7
Cigarette smoking (% current in 1959)	41.5	40.5	45.3	41.9	1.8
Cigarette smoking (% current in 1972)	23.3	24.2	25.9	14.9	1.8
Fruit/fruit juices (7+ times/week)	63.2	60.6	63.7	66.2	59.0

Note. Values in 1959 and 1999 for male subjects in 11 counties who responded to 1999 questionnaire.

TABLE 2
Demographic and lifestyle characteristics in 1959 for California CPS I female subjects as of 1/1/1973 who resided in the 11 counties having 1979–1983 $PM_{2.5}$ measurements and provided 1972 cigarette smoking status

Characteristic	1959 value (11 $PM_{2.5}$ counties)	1959 value (2 highest $PM_{2.5}$ counties)	1959 value (2 lowest $PM_{2.5}$ counties)	1959 value for 1999 respondents	1999 value for 1999 respondents
Mean level of 1979–1983 $PM_{2.5}$ ($\mu g/m^3$)	23.4	36.1	13.1		
Number of subjects alive or lost as of 1/1/1973	22,228	1491	1313		
Lost to follow-up since 1/1/1973 (%)	9.058	9.276	10.252		
Number of subjects alive 1/1/1973 and not lost since 1/1/1973	20,215	1353	1178	2877 (alive 1999)	2877 (alive 1999)
Age as of 1/1/1973 (mean, years)	64.9	66.3	64.0	57.1	57.1
Age as of 1/1/1983 (mean, years)	72.9	74.3	72.0	67.1	67.1
Race (% white)	98.3	99.3	97.9	98.1	97.7
Marital status (% married)	83.1	81.8	86.3	90.0	32.3
Education (% ≥12 yr)	76.8	76.6	82.6	89.9	93.3
Height (mean, inches)	63.8	63.8	64.0	64.1	63.5
Weight (mean, pounds)	137.1	138.4	135.3	133.0	137.4
History of serious diseases (% yes)	9.9	10.6	9.8	5.5	
Cancer	5.9	7.0	6.0	3.8	36.8
Heart disease	3.5	3.0	3.1	1.5	
Stroke	0.5	0.6	0.7	0.2	
Sick at the present time (% yes)	8.4	8.1	5.6	7.2	22.0
Occupation (% professional)	15.9	21.4	16.0	18.6	
Residence location (% urban)	97.7	99.1	98.0	96.9	
Exercise (% moderate or heavy)	80.2	82.8	83.3	77.9	63.4
Cigarette smoking (% current in 1959)	32.5	28.7	40.4	31.1	3.6
Cigarette smoking (% current in 1972)	22.8	19.6	29.1	19.8	3.6
Fruit/fruit juices (7+ times/week)	74.1	75.4	74.3	74.8	60.1

Note. Values in 1959 and 1999 for female subjects in 11 counties who responded to 1999 questionnaire.

26–29.99, >30 kg/m^2), male occupational exposure (no, yes), exercise (none/slight, moderate, heavy), and fruit/fruit juice intake (0, 1, 2, 3, 4, 5, 6, 7 days/wk). One additional variable, health status at entry (good, fair, poor, ill, sick/cancer/CHD/stroke), was evaluated in a sensitivity analysis. The confounding variables are defined at original entry into study in late 1959, except for cigarette smoking status, which was updated in late 1972. All of the confounding variables were measured again in a large sample of survivors during 1999.

Subgroup analyses were done by sex, year of birth (1873–1907, 1908–1929, representing ages 43–64 and 65–99 as of January 1, 1973), education level (<12, 12, 12+ yr), cigarette smoking status (never, former, current as of October 1, 1972), and health status (healthy, unhealthy as of October 1, 1959), as well as by decade of follow-up (January 1, 1973–December 31, 1982, January 1, 1983–December 31, 1992, January 1, 1993–December 31, 2002). In addition, the relative mortality rates by county of residence were calculated using PHREG as an alternative method to assess the influence of different county-wide pollution levels. The Los Angeles county subjects are used as the referent group in estimating the fully-adjusted RRs during 1973–2002, 1973–1982, and 1983–2002 for each of the other 24 counties.

RESULTS

Demographic Characteristics and 1979–1983 $PM_{2.5}$ Data

The late 1959 demographic and lifestyle characteristics of the CA CPS I subjects in the 11 counties with 1979–1983 $PM_{2.5}$ data (mean of 23.4 $\mu g/m^3$) are shown in Table 1 for 15,574 males and in Table 2 for 20,218 females. These tables also show the corresponding characteristics for the subjects in the two counties (Kern and Riverside) with the highest $PM_{2.5}$ levels (mean of 36.1 $\mu g/m^3$) and in the two counties (Contra Costa and Santa Barbara) with the lowest $PM_{2.5}$ levels (mean of 13.1 $\mu g/m^3$). The characteristics of subjects are quite similar, irrespective of their mean pollution levels. The mean age of the subjects alive as of January 1, 1973, was 65.7 yr for males and 64.9 yr for

TABLE 3

Fully adjusted relative risk of death from all causes (RR and 95% CI) during 1/1/1973–12/31/2002 and 1/1/1983–12/31/2002 for both sexes, by county of residence relative to Los Angeles county for the 25 counties with the most California CPS I subjects, based on 10/1/1972 county of residence

County of residence as of 10/1/1972	1973–2002 Deaths/ 1973 subjects	Percent alive or dead same co in 1999	Fully adjusted 1973–2002 RR(95% CI)	Fully adjusted 1973–1982 RR(95% CI)	Fully adjusted 1983–2002 RR(95% CI)
Alameda	3380/4294	60.7	0.962 (0.926–0.999)	0.948 (0.886–1.015)	0.967 (0.924–1.012)
Butte	462/534	73.1	0.999 (0.910–1.096)	0.899 (0.763–1.060)	1.051 (0.939–1.176)
Contra Costa	1260/1652	60.1	0.999 (0.943–1.058)	0.989 (0.890–1.100)	1.004 (0.937–1.076)
Fresno	840/1085	80.0	0.935 (0.872–1.002)	0.896 (0.786–1.021)	0.951 (0.876–1.033)
Humboldt	424/507	79.3	0.992 (0.900–1.092)	0.985 (0.830–1.168)	0.992 (0.882–1.115)
Kern	630/790	79.5	0.944 (0.872–1.023)	0.950 (0.824–1.096)	0.941 (0.854–1.036)
Marin	641/805	57.2	0.939 (0.867–1.016)	1.006 (0.875–1.158)	0.908 (0.825–1.000)
Napa	500/627	73.4	0.949 (0.868–1.038)	0.817 (0.687–0.972)	1.006 (0.906–1.117)
Orange	2453/3050	65.1	0.990 (0.948–1.034)	0.962 (0.891–1.038)	1.003 (0.952–1.056)
Riverside	1311/1575	59.6	0.959 (0.906–1.015)	1.022 (0.928–1.064)	0.926 (0.863–0.993)
Sacramento	1370/1721	77.2	0.998 (0.944–1.055)	0.960 (0.867–1.064)	1.013 (0.948–1.083)
San Bernardino	1340/1622	63.5	0.992 (0.938–1.049)	0.932 (0.841–1.033)	1.018 (0.951–1.088)
San Diego	2958/3615	84.4	0.992 (0.954–1.033)	0.911 (0.847–0.979)	1.028 (0.981–1.078)
San Francisco	1597/2043	48.3	0.963 (0.914–1.014)	0.985 (0.899–1.080)	0.952 (0.894–1.014)
San Joaquin	248/325	71.9	0.925 (0.816–1.049)	0.847 (0.670–1.071)	0.965 (0.832–1.120)
San Mateo	1403/1789	58.0	0.949 (0.899–1.003)	0.897 (0.808–0.997)	0.971 (0.910–1.035)
Santa Barbara	411/516	67.4	0.968 (0.878–1.068)	0.832 (0.690–1.003)	1.030 (0.918–1.156)
Santa Clara	1851/2345	63.5	0.955 (0.910–1.003)	0.961 (0.880–1.049)	0.954 (0.900–1.012)
Santa Cruz	295/372	64.7	0.890 (0.793–0.999)	0.980 (0.805–1.193)	0.849 (0.736–0.979)
Solano	402/505	59.8	0.901 (0.815–0.995)	0.823 (0.685–0.989)	0.934 (0.830–1.051)
Sonoma	482/581	75.7	0.968 (0.884–1.060)	0.919 (0.781–1.082)	0.987 (0.885–1.102)
Stanislaus	551/691	83.7	0.984 (0.904–1.072)	0.981 (0.841–1.144)	0.981 (0.885–1.087)
Tulare	921/1117	78.7	1.047 (0.979–1.119)	1.031 (0.918–1.158)	1.054 (0.972–1.144)
Ventura	369/474	69.1	0.967 (0.872–1.072)	0.774 (0.629–0.951)	1.053 (0.935–1.187)
Los Angeles	13,747/17,340	64.4	1.000	1.000	1.000
Total	39,846/49,975	66.4			
Chi-square test of homogeneity (24 degrees of freedom)			$\chi^2 = 27.48$ $p = .283$	$\chi^2 = 32.21$ $p = .122$	$\chi^2 = 30.70$ $p = .163$

TABLE 4

Fully adjusted relative risk of death from all causes (RR and 95% CI) by county of residence relative to Los Angeles county, during 1973–2002 and 1983–2002 for both sexes, for the 35,789 California CPS I subjects in rank order of 1979–1983 $PM_{2.5}$ level for the 11 counties and groups of these counties

County of residence as of 10/1/1972	1/1/1973 subjects	Fully adjusted 1973–2002 RR(95% CI)	Fully adjusted 1973–1982 RR(95% CI)	Fully adjusted 1983–2002 RR(95% CI)	$PM_{2.5}$ ($\mu g/m^3$) 1979–1983	Rank
Santa Barbara	516	0.968 (0.878–1.068)	0.832 (0.690–1.003)	1.030 (0.918–1.156)	10.6	1
Contra Costa	1652	0.999 (0.943–1.058)	0.989 (0.890–1.100)	1.004 (0.937–1.076)	13.9	2
Alameda	4294	0.962 (0.926–0.999)	0.948 (0.886–1.015)	0.967 (0.924–1.012)	14.4	3
Butte	534	0.999 (0.910–1.096)	0.899 (0.763–1.060)	1.051 (0.939–1.176)	15.5	4
San Francisco	2043	0.963 (0.914–1.014)	0.985 (0.899–1.080)	0.952 (0.894–1.014)	16.4	5
Santa Clara	2345	0.955 (0.910–1.003)	0.961 (0.880–1.049)	0.954 (0.900–1.012)	17.8	6
Fresno	1085	0.935 (0.872–1.002)	0.896 (0.786–1.021)	0.951 (0.876–1.033)	18.4	7
San Diego	3615	0.992 (0.954–1.033)	0.911 (0.847–0.979)	1.028 (0.981–1.078)	18.9	8
Los Angeles	17,340	1.000	1.000	1.000	28.2	9
Kern	790	0.944 (0.872–1.023)	0.950 (0.824–1.096)	0.941 (0.854–1.036)	30.9	10
Riverside	1575	0.959 (0.906–1.015)	1.022 (0.928–1.064)	0.926 (0.863–0.993)	42.0	11
Total	35,789				23.4	
Chi-square test of homogeneity (10 degrees of freedom)		$\chi^2 = 12.65$ $p = .244$	$\chi^2 = 14.27$ $p = .161$	$\chi^2 = 16.35$ $p = .090$		
Two lowest $PM_{2.5}$ counties (Contra Costa and Santa Barbara)		0.991 (0.942–1.043)	0.948 (0.863–1.041)	1.010 (0.951–1.073)	13.1	
Next lowest $PM_{2.5}$ counties (Alameda, Butte, San Francisco)		0.965 (0.935–0.996)	0.955 (0.903–1.010)	0.969 (0.933–1.006)	15.1	
Next highest $PM_{2.5}$ counties (Fresno, San Diego, Santa Clara)		0.971 (0.942–1.002)	0.926 (0.875–0.980)	0.991 (0.955–1.029)	18.5	
Reference county (Los Angeles)		1.000	1.000	1.000	28.2	
Two highest $PM_{2.5}$ counties (Kern and Riverside)		0.954 (0.910–1.001)	1.000 (0.921–1.087)	0.931 (0.878–0.986)	36.1	
Chi-square test of homogeneity (4 degrees of freedom)		$\chi^2 = 8.48$ $p = .075$	$\chi^2 = 8.72$ $p = .069$	$\chi^2 = 8.02$ $p = .091$		

females and their minimum age was 43 yr. The mean age of the subjects alive as of January 1, 1983, was 73.8 yr for males and 72.9 yr for females, and their minimum age was 53 yr.

The 1979–1983 $PM_{2.5}$ data from the IPN are shown in Appendix Table 1 for 11 California counties, with details for the 15 monitoring sites at which measurements were made.

The 1999 follow-up questionnaire provided important information about the confounding variables for survivors 40 yr after they originally enrolled in the study. Although these survivors were the younger members of the cohort, with a mean age of about 57 yr as of January 1, 1973, they provide a good indication of the risk factor changes that have occurred. A comparison of their 1959 and 1999 responses in Tables 1 and 2 shows that the variables of race, education level, exercise, body mass index, and fruit/fruit juice intake changed very little and were similar in the high and low $PM_{2.5}$ counties. The percentage of married subjects declined substantially in all counties because of the large fraction of spouses who died. The percentage of current cigarette smokers declined dramatically and uniformly in all counties, reflecting the large degree of smoking cessation that has already been documented in this cohort (Enstrom & Heath, 1999). Health status, used as an additional variable in a sensitivity analysis, also declined substantially among subjects in all counties, because the aging survivors had a much higher prevalence of cancer and other diseases in 1999 than they did in 1959.

Relative Risks by County of Residence

Table 3 shows the 1973–2002, 1973–1982, and 1983–2002 mortality risks relative to Los Angeles county, adjusted for age, sex, and eight confounding variables, for the 25 counties with the most CA CPS I subjects, including the 11 counties with 1979–1983 $PM_{2.5}$ data. Overall, the RRs were quite consistent with each other and most had a 95% CI that included 1.0. None of the RRs were greatly different from 1.0. Wald chi-square tests conducted on these RRs did not reject the hypothesis of homogeneity during any of the three follow-up periods. Also, Table 3 shows that, based on their counties of residence and death from 1972 to 1999, about 66% of the subjects remained in the same county during this period, indicating relative stability of residence. In particular, the stability of residence was similar in the two highest $PM_{2.5}$ counties (66%), the two lowest $PM_{2.5}$ counties (62%), and in Los Angeles county (64%).

Table 4 shows the mortality risks relative to Los Angeles county for 11 counties ranked in order by their 1979–1983 $PM_{2.5}$ value. During 1973–2002, the two counties (Kern and Riverside) with highest $PM_{2.5}$ levels (mean of 36.1 $\mu g/m^3$) had an RR of 0.954 (0.910–1.001), whereas the two counties (Contra Costa and Santa Barbara) with the lowest $PM_{2.5}$ levels (mean of 13.1 $\mu g/m^3$) had a slightly higher RR of 0.991 (0.942–1.043). During 1983–2002 there was a larger difference, with corresponding RRs of 0.931 (0.878–0.986) and 1.010 (0.951–1.073). During 1973–1982 there was a reverse pattern, with corresponding RRs of 1.000 (0.921–1.087) and 0.948 (0.863–1.041). Two groups of the six counties with medium $PM_{2.5}$ levels had intermediate RRs that were consistent with the RRs for the high and low $PM_{2.5}$ counties. Although there is some variation, Wald chi-square tests conducted on these RRs did not reject the hypothesis of homogeneity during any of the three follow-up periods for the individual counties or for the groups of counties. These findings are consistent with those in Table 3.

Relative Risks by 1979–1983 $PM_{2.5}$ Level

Table 5 shows the relationship of 1973–2002 mortality to 1979–1983 $PM_{2.5}$ level, by decade of follow-up, based on assigning each subject the $PM_{2.5}$ level of the county in which they resided as of late 1972. Both the age- and sex-adjusted and fully adjusted 1973–2002 RRs are shown. Also shown are the 1973–2002 RRs for selected subgroups defined by sex, age (year of birth), education level, cigarette smoking status, and health status. These RRs were calculated based on a unit increase in $PM_{2.5}$ of 10 $\mu g/m^3$. The age- and sex-adjusted RRs and the fully adjusted RRs were slightly elevated (~1.04) during the first decade, 1973–1982, but were essentially 1.0 during the next two decades. Among the subgroups, the fully adjusted 1973–2002 RRs were slightly elevated (1.03) for females and younger subjects (those born during 1908–1929) and consistent with 1.0 for the others.

Table 6 shows the relationship of 1973–1982 mortality to 1979–1983 $PM_{2.5}$ levels for all subjects and for the same selected subgroups. The fully adjusted RRs were significantly elevated above 1.0 for all subjects (1.04), for females (1.05), for younger subjects (1.06), for the least educated (1.07), for former smokers (1.06), and for the healthy (1.05). Table 7 shows the relationship of 1983–2002 mortality to 1979–1983 $PM_{2.5}$ levels for all subjects and for the subgroups. Results for the entire 20 years are shown because they are virtually the same as the separate results for 1983–1992 and 1993–2002. The fully adjusted RRs were not elevated above 1.0 for any subgroup during 1983–2002. Taken as a whole, these results suggest there was a weak relationship between fine particulate pollution and mortality during 1973–1982, but none during 1983–2002. However, because of statistical fluctuation, effects of up to a 2% increase in mortality per 10 $\mu g/m^3$ of $PM_{2.5}$ cannot be ruled out during 1983–2002.

Sensitivity Analysis of Relative Risks

Table 8 presents a sensitivity analysis to determine the extent to which the RRs in Tables 5–7 were influenced by the confounding variables that were used. It shows the RRs during 1973–2002, 1973–1982, and 1983–2002 based on sequential proportional hazards regression models, beginning with 1979–1983 $PM_{2.5}$ as the only independent variable and then adding age, sex, and nine confounding variables, one at a time. The results were not particularly sensitive to the addition of any of the variables except age and cigarette smoking status. It is unknown how much the RRs would be changed if $PM_{2.5}$ data were available for years before and after 1979–1983. None of the RRs during 1973–2002 and 1983–2002 were significantly elevated above 1.0 after adjustment for age, but the RRs have

TABLE 5

Relative risk of death from all causes (RR and 95% CI) during 1/1/1973–12/31/2002 associated with change of 10 $\mu g/m^3$ in 1979–1983 $PM_{2.5}$, with subgroups defined by decade of follow-up, sex, year of birth, education level, cigarette smoking status as of 10/1/1972, and health status as of 10/1/1959

		Model based on 1979–1983 $PM_{2.5}$	
Subgroups	Deaths/subjects	Age-sex-adjusted RR (95% CI)	Fully adjusted RR (95% CI)
All subjects during each decade of full follow-up period (1/1/1973–12/31/2002)			
1/1/1973–12/31/1982	8795/35,783	1.032 (1.003–1.062)	1.039 (1.010–1.069)
1/1/1983–12/31/1992	10,821/26,988	0.989 (0.964–1.015)	0.996 (0.970–1.022)
1/1/1993–12/31/2002	8825/16,167	0.997 (0.969–1.026)	0.999 (0.970–1.028)
Subjects during full follow-up period (1/1/1973–12/31/2002)			
All subjects	28,441/35,783	1.005 (0.989–1.021)	1.010 (0.994–1.026)
All males	13,532/15,573	0.996 (0.973–1.020)	0.993 (0.970–1.016)
All females	14,909/20,210	1.013 (0.991–1.035)	1.027 (1.005–1.050)
Born 1908–1929 (1973 age 43–64)	13,354/20,086	1.022 (0.998–1.046)	1.027 (1.003–1.052)
Born 1873–1907 (1973 age 65–99)	15,082/15,697	0.991 (0.970–1.013)	0.996 (0.975–1.018)
<12 yr education	8025/9079	1.016 (0.987–1.047)	1.018 (0.989–1.049)
12 yr education	6346/8557	1.002 (0.969–1.037)	1.005 (0.972–1.040)
>12 yr education	14,070/18,147	0.998 (0.976–1.021)	1.007 (0.984–1.030)
Never smoker	11,528/15,181	1.020 (0.995–1.045)	1.019 (0.994–1.044)
Former smoker	10,074/12,400	1.005 (0.978–1.032)	1.005 (0.978–1.032)
Current smoker	6839/8202	1.003 (0.971–1.036)	0.999 (0.967–1.032)
Healthy	22,234/28,461	1.006 (0.988–1.024)	1.010 (0.992–1.028)
Unhealthy	5456/6439	0.981 (0.945–1.018)	0.993 (0.957–1.030)

TABLE 6

Relative risk of death from all causes (RR and 95% CI) during 1/1/1973–12/31/1982 associated with change of 10 $\mu g/m^3$ in 1979–1983 $PM_{2.5}$, with subgroups defined by sex, year of birth, education level, cigarette smoking status as of 10/1/1972, and health status as of 10/1/1959

		Model based on 1979–1983 $PM_{2.5}$	
Subgroups	Deaths/subjects	Age-sex-adjusted RR (95% CI)	Fully adjusted RR (95% CI)
Subjects during initial decade of follow-up (1/1/1973–12/31/1982)			
All subjects	8795/35,783	1.032 (1.003–1.062)	1.039 (1.010–1.069)
All males	4701/15,573	1.027 (0.988–1.067)	1.029 (0.990–1.069)
All females	4094/20,210	1.039 (0.996–1.083)	1.052 (1.009–1.096)
Born 1908–1929 (1973 age 43–64)	2637/20,086	1.062 (1.006–1.120)	1.064 (1.008–1.122)
Born 1873–1907 (1973 age 65–99)	6158/15,697	1.021 (0.988–1.056)	1.031 (0.997–1.066)
<12 yr education	3123/9079	1.064 (1.015–1.115)	1.072 (1.023–1.124)
12 yr education	1686/8557	1.045 (0.980–1.115)	1.046 (0.981–1.116)
>12 yr education	3986/18,147	1.001 (0.960–1.045)	1.011 (0.969–1.055)
Never smoker	3425/15,181	1.038 (0.993–1.086)	1.038 (0.992–1.085)
Former smoker	3264/12,400	1.059 (1.011–1.110)	1.058 (1.010–1.109)
Current smoker	2106/8202	1.014 (0.957–1.075)	1.009 (0.952–1.069)
Healthy	6432/28,461	1.043 (1.009–1.078)	1.050 (1.016–1.085)
Unhealthy	2104/6439	0.981 (0.925–1.040)	0.991 (0.935–1.051)

TABLE 7
Relative risk of death from all causes (RR and 95% CI) during 1/1/1983–12/31/2002 associated with change of 10 $\mu g/m^3$ in 1979–1983 $PM_{2.5}$, with subgroups defined by sex, year of birth, education level, cigarette smoking status as of 10/1/1972, and health status as of 10/1/1959

		Model based on 1979–1983 $PM_{2.5}$	
Subgroups	Deaths/subjects	Age-sex-adjusted RR (95% CI)	Fully adjusted RR (95% CI)
Subjects during last two decades of follow-up (1/1/1983–12/31/2002)			
All subjects	19,646/26,988	0.992 (0.973–1.011)	0.997 (0.978–1.016)
All males	8831/10,872	0.979 (0.951–1.008)	0.974 (0.947–1.003)
All females	10,815/16,116	1.003 (0.978–1.029)	1.018 (0.992–1.044)
Born 1908–1929 (1983 age 53–74)	10,717/17,449	1.012 (0.986–1.040)	1.018 (0.992–1.046)
Born 1873–1907 (1983 age 75–99)	8929/9539	0.972 (0.945–0.999)	0.975 (0.948–1.002)
<12 yr education	4902/5956	0.987 (0.950–1.025)	0.986 (0.949–1.024)
12 yr education	4660/6871	0.985 (0.947–1.026)	0.990 (0.951–1.030)
>12 yr education	10,084/14,161	0.997 (0.970–1.024)	1.005 (0.978–1.032)
Never smoker	8103/11,756	1.011 (0.982–1.042)	1.011 (0.981–1.041)
Former smoker	6810/9136	0.979 (0.947–1.012)	0.980 (0.949–1.013)
Current smoker	4733/6096	0.999 (0.961–1.039)	0.996 (0.958–1.036)
Healthy	15802/22,029	0.990 (0.969–1.012)	0.994 (0.973–1.015)
Unhealthy	3352/4335	0.980 (0.935–1.028)	0.995 (0.949–1.043)

upper confidence intervals as high as 1.028. All of the RRs during 1973–1982 were significantly elevated about 1.0, with upper confidence intervals as high as 1.071. Based on their large χ^2 values, the variables of age, sex, and cigarette smoking status were by far the most important variables in the model. After inclusion of age, sex, and cigarette smoking status, the addition of the next 7 independent variables changed the RRs by only about 0.1%. When initial health status was entered as an additional

TABLE 8
Relative risk of death from all causes (RR and 95% CI) during 1/1/1973–12/31/2002, 1/1/1973–12/31/1982, and 1/1/1983–12/31/2002 associated with change of 10 $\mu g/m^3$ in 1979–1983 $PM_{2.5}$, by individual confounding variables defined as of 1959, except for 1972 cigarette smoking; Age, sex, and nine confounding variables are added to the proportional hazards regression model one variable at a time

Cumulative PHREG model based on adding one variable at a time	1973–2002 Chi-square	1973–2002 RR (95% CI)	1973–1982 RR (95% CI)	1983–2002 RR (95% CI)
1979–1983 $PM_{2.5}$	0.68	1.029 (1.012–1.045)	1.058 (1.028–1.089)	1.016 (0.996–1.035)
+Age	14,445.92	1.003 (0.987–1.019)	1.029 (1.000–1.059)	0.991 (0.972–1.010)
+Sex	464.16	1.005 (0.989–1.021)[a]	1.032 (1.003–1.062)[a]	0.992 (0.973–1.011)[a]
+Cigarette smoking	1,610.59	1.011 (0.995–1.028)	1.041 (1.012–1.071)	0.998 (0.979–1.017)
+Race	0.93	1.011 (0.995–1.027)	1.041 (1.012–1.071)	0.997 (0.978–1.017)
+Education	53.67	1.011 (0.995–1.027)	1.041 (1.012–1.071)	0.997 (0.978–1.016)
+Martial status	26.16	1.011 (0.995–1.027)	1.041 (1.012–1.071)	0.997 (0.978–1.016)
+Body mass index	91.87	1.010 (0.994–1.026)	1.040 (1.011–1.070)	0.996 (0.977–1.016)
+Occupational exposure	3.28	1.010 (0.994–1.026)	1.040 (1.011–1.070)	0.996 (0.977–1.016)
+Exercise	0.02	1.010 (0.994–1.026)	1.039 (1.010–1.069)	0.997 (0.978–1.016)
+Fruit/fruit juice intake	42.54	1.010 (0.994–1.026)[b]	1.039 (1.010–1.069)[b]	0.997 (0.978–1.016)[b]
+Health status	156.15	1.007 (0.991–1.023)	1.036 (1.006–1.066)	0.994 (0.975–1.013)

[a] Age-sex-adjusted RR.
[b] Eight-variable fully adjusted RR.

TABLE 9

Fully-adjusted relative risk of death from all causes (RR and 95% CI) by cigarette smoking status as of 10/1/1972, during 1973–2002, 1973–1982, and 1983–2002 for both sexes, for the California CPS subjects in 11 counties with 1979–1983 $PM_{2.5}$ measurements, which were used as one of the confounding variables

Cigarette smoking status as of 10/1/1972	Fully adjusted 1973–2002 RR (95% CI)	Fully adjusted 1973–1982 RR (95% CI)	Fully adjusted 1983–2002 RR (95% CI)
Deaths/subjects	28,447/35,789	8801/35,789	19,646/26,988
Never (as of 1959 and 1972)	1.000	1.000	1.000
Former (as of 1959 and 1972)	1.054 (1.014–1.096)	1.061 (0.987–1.140)	1.049 (1.001–1.100)
Former (as of 1972 only)	1.253 (1.212–1.295)	1.312 (1.236–1.392)	1.224 (1.176–1.273)
Current: 1–9 cpd (as of 1972)	1.239 (1.150–1.336)	1.227 (1.065–1.414)	1.243 (1.138–1.357)
Current: 10–19 cpd (as of 1972)	1.597 (1.510–1.688)	1.667 (1.508–1.842)	1.566 (1.465–1.675)
Current: 20 cpd (as of 1972)	1.871 (1.791–1.953)	1.829 (1.689–1.980)	1.887 (1.792–1.987)
Current: 21–39 cpd (as of 1972)	2.068 (1.936–2.210)	1.889 (1.666–2.140)	2.145 (1.984–2.320)
Current: 40+ cpd (as of 1972)	2.543 (2.375–2.723)	2.460 (2.189–2.765)	2.587 (2.378–2.814)
Chi-square test of homogeneity (7 degrees of freedom)	$\chi^2 = 1701.75$ $p < .0001$	$\chi^2 = 478.14$ $p < .0001$	$\chi^2 = 1232.17$ $p < .0001$

Note. cpd, cigarettes per day.

independent variable, the RRs decreased by 0.3%. However, initial health status was not one of the eight confounding variables used in calculating the fully adjusted RRs in the other tables, because it may have been influenced by exposure to air pollution before entry into the study.

Relative Risks by 1972 Cigarette Smoking Level

Table 9 shows the fully adjusted RRs for eight levels of 1972 cigarette smoking, the strongest confounding variable in this study. Note there was a strong and clear dose-response relationship during 1973–2002. The dose-response relationship remained as strong during 1983–2002 as it was during 1973–1982, in spite of the large degree of smoking cessation that occurred from 1972 to 1999, as documented in Tables 1 and 2. This comparison supports the findings of an earlier paper, which examined smoking cessation and mortality trends in this cohort during 1960–1997 (Enstrom & Heath, 1999). The Wald chi-square test of homogeneity for each of the three follow-up periods, where $\chi^2 > 478$ for 7 degrees of freedom, clearly rejects the hypothesis that the RRs are equal (homogeneous). The large RRs related to increasing cigarette smoking level are shown here in order to put the RRs related to increasing $PM_{2.5}$ level in Table 4 in perspective.

DISCUSSION

Strengths and Uncertainties of This Study

This study has several important strengths: a large, diverse cohort of males and females distributed throughout California, a large number of deaths, extensive baseline and follow-up data on demographic and lifestyle characteristics, long-term follow-up of a high percentage of subjects, relative stability of subjects based on their residential address history, and availability of $PM_{2.5}$ measurements for over 70% of the subjects. In addition, there is a wide range of $PM_{2.5}$ levels (10.6 to 42.0 $\mu g/m^3$) available for subjects in 11 counties. Although the CA CPS I cohort is not a random sample of the California population, previous examination has shown that mortality ratios based on cigarette smoking status are similar in this cohort (Enstrom & Heath, 1999) and in a cohort representative of the US population (Enstrom, 1999).

The results of this study, as in all epidemiology studies, are dependent upon the underlying data and the analytical methods that were used. Major uncertainties include the extent to which the available air quality data represent actual exposures, the validity of the proportional hazards regression calculations of the RRs, and the potentially important confounders that may have been omitted from the analysis.

The $PM_{2.5}$ air quality data are limited to 11 California counties and 8 of these counties had only one outdoor monitoring station each. The period of monitoring was from July 1979 to December 1983 and most counties had data for just a two to three year period. These sample data are assumed to represent long-term personal exposures of each subject based on their county of residence in late 1972. The validity of this assumption has not been confirmed, but these same limited data have been used in other major cohort studies (Pope et al., 1995, 2002; Lipfert et al., 2000, 2003).

The assumption that individual exposures are the same as county-wide averages, as measured by a few centrally-located monitors, can result in the "ecological fallacy," where results based on group averages differ from those based on individual exposures (Piantadosi et al., 1988). However, it is impractical to monitor individual exposures for a large cohort, especially

TABLE 10

Relative risk (RR) and 95% confidence interval (CI) for long-term all-cause mortality per 10-$\mu g/m^3$ increase in $PM_{2.5}$ for U.S. cohort studies based on $PM_{2.5}$ data, circa 1980

	$PM_{2.5}$		Study characteristics					
Study (author, year)	Data period	Mean (range) ($\mu g/m^3$)	Cohort geographic definition	Follow-up period	Mean entry age for period	Number entered in cohort	Deaths in follow-up period	RR (95% CI)
Males								
Dockery et al., 1993	1979–1985	19 (11–30)	6 U.S. cities	1975–1989	~50	3671[a]	830[a]	1.15 (1.02–1.30)[b]
Pope et al., 1995	1979–1981	18 (9–34)	50 U.S. SMSAs	1982–1989	57	130,310[a]	~12,400[a]	1.07 (1.03–1.11)[b]
McDonnell et al., 2000	1973–1977	32 (17–45)	9 CA airsheds	1976–1992	58	≤1347	≤375	1.09 (0.98–1.21)[b]
Lipfert et al., 2000	1979–1981	24 (6–42)	42 U.S. counties	1975–1981	51	26,067	~4600[c]	0.95 (0.89–1.01)[c]
	1982–1984	22 (8–41)		1982–1988	57	~21,467	~6100[c]	0.94 (0.90–0.98)[c]
	1982–1984	22 (8–41)		1989–1996	63	~15,367	~5765[c]	0.89 (0.85–0.95)[c]
Pope et al., 2002	1979–1983	21 (10–30)	61 U.S. SMSAs	1982–1998	57	~159,000[a]	~36,000[a]	1.05 (1.01–1.10)
Enstrom, 2005	1979–1983	24 (11–42)	11 CA counties	1973–1982	66	15,573	4701	1.03 (0.99–1.07)
	1979–1983	24 (11–42)		1983–2002	74	10,872	8831	0.97 (0.95–1.00)
Females								
Dockery et al., 1993	1979–1985	19 (11–30)	6 U.S. cities	1975–1989	~50	4440[a]	599[a]	1.12 (0.96–1.30)[b]
Pope et al., 1995	1979–1981	18 (9–34)	50 U.S. SMSAs	1982–1989	57	164,913[a]	~8365[a]	1.06 (1.01–1.12)[b]
McDonnell et al., 2000	1973–1977	32 (17–45)	9 CA airsheds	1976–1992	58	≤2422	≤568	~1.00 (assumed)
Pope et al., 2002	1979–1983	21 (10–30)	61 U.S. SMSAs	1982–1998	57	~200,000[a]	~24,000[a]	1.02 (0.98–1.06)
Enstrom, 2005	1979–1983	24 (11–42)	11 CA counties	1973–1982	65	20,210	4094	1.05 (1.01–1.10)
	1979–1983	24 (11–42)		1983–2002	73	16,116	10,815	1.02 (0.99–1.04)
Both Sexes								
Dockery et al., 1993	1979–1985	19 (11–30)	6 U.S. cities	1975–1989	~50	8111	1430	1.13 (1.04–1.23)[b]
Pope et al., 1995	1979–1981	18 (9–34)	50 U.S. SMSAs	1982–1989	57	295,223	20,765	1.07 (1.04–1.10)[b]
Pope et al., 2002	1979–1983	21 (10–30)	61 U.S. SMSAs	1982–1998	57	~359,000	~60,000	1.04 (1.01–1.08)
Enstrom, 2005	1979–1983	24 (11–42)	11 CA counties	1973–1982	65	35,783	8795	1.04 (1.01–1.07)
	1979–1983	24 (11–42)		1983–2002	73	26,988	19,646	1.00 (0.98–1.02)

[a]Obtained from supplementary data (Krewski et al., 2000).

[b]Recalculated from published data (US EPA, 2004).

[c]Obtained from the author.

over the long-term. This analysis used the smallest practical geographic unit (counties), given typical mobility of the subjects, in hopes of minimizing exposure errors.

The relative risk results are all based on Cox proportional hazards regression (PHREG), which has been used in numerous cohort survival analyses. This statistical methodology depends on the assumption of proportionality (SAS, 2004). The validity of this assumption with respect to the $PM_{2.5}$ variable has been confirmed for most RRs in Tables 3–9 by the Kolmogorov-type supremum test of functional form (SAS, 2004). Also, previous findings on cigarette smoking and mortality in this cohort showed that relative risks based on proportional hazards regression were similar to relative risks based on life table survival analysis (Enstrom & Heath, 1999).

With respect to the impact of additional potential confounding variables, ecological variables at the county level, such as climate, were explored and found to be uncritical. The analyses based on the eight-variable model found that after age and sex adjustment only one confounding variable, cigarette smoking status, had much impact on the RRs. Indeed, each fully-adjusted RR was within 1.5% of the corresponding age-sex-adjusted RR.

Comparisons with Other Cohort Studies

In Table 10 the major findings in this study are compared with those of the other U.S. cohort studies based on $PM_{2.5}$ data, circa 1980. The basic characteristics and relative risks for all-cause mortality are given for all six studies, where each relative risk (RR and 95% CI) is based on a 10 $\mu g/m^3$ increase in $PM_{2.5}$. The RRs were standardized to the extent possible by using the same conversion of published results that was used by the EPA in their latest criteria document on particulate matter (US EPA, 2004). For instance, the RR of 1.07 (1.04–1.10) during 1982–1989 for both sexes in the ACS CPS II study (Pope et al., 1995) is based on a mortality difference of 17% between the highest and lowest $PM_{2.5}$ areas and a $PM_{2.5}$ difference of 25 $\mu g/m^3$. By comparison, the RR of 1.00 (0.98–1.02) during 1983–2002 among both sexes in the current CA CPS I study is based on a $PM_{2.5}$ variation up to 31 $\mu g/m^3$ among 11 counties.

The RRs in Table 10 range from 0.89 to 1.15 and each corresponding 95% CI either includes 1.0 or is within 0.05 of including 1.0. Thus, the relationship between $PM_{2.5}$ and mortality is very weak and near the limit of detectability by epidemiologic methods. In order to define this relationship as accurately as possible, it is important to understand the differences that exist between the RRs. These differences could be due to the epidemiologic methodology used or they could be due to the characteristics of the study cohorts, such as their geographic location, follow-up period, demographics, and size. For instance, the RR of 1.07 in the ACS CPS II study is the average relationship between $PM_{2.5}$ level and mortality in 50 areas of the U.S. However, detailed reanalysis of this study reveals substantial geographic variation in the relationship (Krewski et al., 2000). In particular, a map of $PM_{2.5}$ levels and relative risk of mortality throughout the U.S. (Figure 21) shows that most areas of California had medium mortality risk and no areas had high mortality risk. This pattern suggests that the relationship between $PM_{2.5}$ and mortality among the California subjects in the CPS II cohort was weaker than the RR of 1.07 and consistent with the RR of 1.00 found in the CA CPS I cohort.

It is clear from Table 10 that no single result can adequately describe the relationship between $PM_{2.5}$ and mortality for the entire country. The complete body of epidemiologic evidence should be used to assess this relationship as accurately as possible within the limitations of epidemiologic methodology. A full comparative examination of the available cohort studies is warranted. Ideally, a standardized method of analysis should be applied to the underlying data in each cohort and the results should be presented in a standardized way. Such an analysis would make a substantial contribution to the research priorities for particulate matter (National Research Council, 2004).

REFERENCES

California Air Resources Board. 2003. *Ambient air quality standards for suspended particulate matter (PM) and sulfates*. California Environmental Protection Agency, Sacramento, CA, July 5. (ftp://ftp.arb.ca.gov/carbis/regact/aaqspm/isor.pdf and www.arb.ca.gov/regact/aaqspm/aaqspm.htm).

Dockery, D. W., Pope, C. A., III, Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G., and Speizer, F. E. 1993. An association between air pollution and mortality in six U.S. cities. *N. Engl. J. Med.* 329:1753–1759.

Enstrom, J. E. 1999. Smoking cessation and mortality trends among two United States populations. *J Clin Epidemiol* 52:813.

Enstrom, J. E., and Heath, C. W., Jr. 1999. Smoking cessation and mortality trends among 118,000 Californians, 1960–97. *Epidemiology* 10:500–512.

Enstrom, J. E., and Kabat, G. C. 2003. Environmental tobacco smoke and tobacco related mortality in a prospective study of Californians, 1960–98. *Br. Med. J.* 326:1057–1061.

Gamble, J. F. 1998. $PM_{2.5}$ and mortality in long-term prospective cohort studies: Cause-effect or statistical associations. *Environ. Health Prospect.* 106:535–549.

Greenbaum, D. S., Bachmann, J. D., Krewski, D., Samet, J. M., White, R., and Wyzga, R. E. 2001. Particulate air pollution standards and morbidity and mortality: Case study. *Am. J. Epidemiol.* 154:S78–S90.

Hammond, E. C. 1972. Smoking habits and air pollution in relation to lung cancer. In *Environmental Factors in Respiratory Disease* (Edited by DH Lee). Fogerty International Center Proceedings No. 11. New York: Academic Press, 1972, Chapter 12, pages 177–198.

Hammond, E. C., and Garfinkel, L. 1980. General air pollution and cancer in the United States. *Prev. Med.* 9:206–211.

Hinton, D. O., Sune, J. M., Suggs, J. C., and Barnard, W. F. 1984. *Inhalable Particulate Network Report: Operation and Data Summary (Mass Concentrations Only)*. Volume I. April 1979–December 1982. EPA-600/4-84-088a. Research Triangle Park, NC: U.S. Environmental Protection Agency, November 1984, particularly pages 108–113.

Hinton, D. O., Sune, J. M., Suggs, J. C., and Barnard, W. F. 1986. *Inhalable Particulate Network Report: Data Summary (Mass*

Concentrations only). Volume III. January 1983-December 1984, EPA-600/4-86/019. Research Triangle Park, NC: U.S. Environmental Protection Agency, April 1986, particularly pages 53–55.

Kaiser, J. 2005. Mounting evidence indicts fine-particle pollution. *Science* 307:1858–1961 (http://www.sciencemag.org/cgi/reprint/307/5717/1858a.pdf).

Krewski, D., Burnett, R. T., Goldberg, M. S., Hoover, K., Siemiatycki, J., Abrahamowicz, M., White, W. H. and others. 2000. *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: Special report.* Cambridge, MA: Health Effects Institute. Part I. Replication and Validation (http://www.healtheffects.org/Pubs/Rean-part1.pdf) and Part II. Sensitivity Analyses (http://www.healtheffects.org/Pubs/Rean-part2.pdf), particularly Figure 21 on p. 197.

Lipfert, F. W. 2003. Commentary on the HEI Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. *J. Toxicol. Environ. Health A* 66:1705–1714.

Lipfert, F. W. 1994. *Air pollution and community health: A critical review and data sourcebook.* New York: van Nostrand Reinhold.

Lipfert, F. W., Malone, R. G., Daum, M. L., Mendell, N. R., and Yang, C. C. 1988. *A statistical study of the macroepidemiology of air pollution and total mortality*. Brookhaven National Laboratory, Upton, NY. Report BNL 52122, April.

Lipfert, F. W., Perry, H. M., Jr., Miller, J. P., Baty, J. D., Wyzga, R. E., and Carmody, S. E. 2000. The Washington University–EPRI veterans' cohort mortality study: Preliminary results. *Inhal. Toxicol.* 12[S4]:41–73.

Lipfert, F. W., Perry, H. M., Jr., Miller, J. P., Baty, J. D., Wyzga, R. E., and Carmody, S. E. 2003. Air pollution, blood pressure, and their long-term associations with mortality. *Inhal. Toxicol.* 15(5):493–512.

Logan, W. P. D., and Glasg M. D. 1953. Mortality in London fog incident, 1952. *Lancet* 1:336–338.

McDonnell, W. F., Nishino-Ishikawa, N., Petersen, F. F., Chen, L. H., and Abbey, D. E. 2000. Relationship of mortality with the fine and coarse fractions of long-term ambient PM_{10} concentrations in nonsmokers. *J. Expos. Anal. Environ. Epidemiol.* 10:427–436.

Moolgavkar, S. H. 1996. A critical review of the evidence on particulate air pollution and mortality. *Epidemiology* 7:420–428.

National Research Council, Committee on Research Priorities for Airborne Particulate Matter 2004. *Research priorities for airborne particulate matter: IV. Continuing research progress.* Washington, DC: National Academies Press. (http://www.nap.edu/books/0309091993/html).

Office of Management and Budget. 2003. *Proposed bulletin on peer review and information quality.* September 15. 68FR 54023-9. (http://www.whitehouse.gov/omb/inforeg/agency_info_quality_links.html).

Ozkaynak, H., and Thurston, G. D. 1987. Associations between 1980 U.S. mortality rates and alternative measures of airborne particle concentration. *Risk Analy.* 7:449–461.

Phalen, R. F. 2002. *The particulate air pollution controversy: A case study and lessons learned.* Boston: Kluger Academic.

Piantadosi, S., Byar, D. P., and Green, S. B. 1988. The ecological fallacy. *Am. J. Epidemiol.* 127:893–904.

Pope, C. A. III, and Dockery, D. W. 1999. Epidemiology of particle effects. In *Air pollution and health*, eds. S. T. Holgate, H. Koren, R. Maynard, and J. Samet, pp. 673–705. London: Academic Press.

Pope, C. A. III, Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E., and Heath, C. W., Jr. 1995. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am. J. Respir. Crit. Care Med.* 151:669–674.

Pope, C. A. III, Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., and Thurston, G. D. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc.* 287:1132–1141.

SAS Institute, Inc. 2004. The PHREG Procedure. *SAS/STAT 9.1 user's guide.* pp. 3213–3332. (http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/stat_ug_7313.pdf). Cary, NC: SAS Institute.

Steinbrook, R. 2004. Peer review and federal regulations. *N. Engl. J. Med.* 350:103–104.

Sune, J. 1999. Personal communication with FW Lipfert. Inhalable Particle Monitoring Network data (*ipmnslrt.exe*). Research Triangle Park, NC: U.S. Environmental Protection Agency.

U.S. Environmental Protection Agency. 1997. Air quality criteria for particulate matter. *Fed. Reg.* 62:38676.

U.S. Environmental Protection Agency. 2004. *Air Quality Criteria for Particulate Matter.* Volume I (EPA/600/P-99/002aF) and Volume II (EPA/600/P-99/002bF). Washington, DC, October 2004, particularly Table 8–12 on page 8–117 (http://cfpub.epa.gov/ncea/cfm/partmatt.cfm).

APPENDIX TABLE 1

1979–1983 $PM_{2.5}$ ("FINE15") data in California from the EPA Inhalable Particulate Network by county and monitoring site (Hinton et al., 1984, pages 108–113; Hinton et al., 1986, pages 53–55; Sune, 1999).

County	City (Street)	SAROAD site	Monitoring period	Samples	Site $PM_{2.5}$ ($\mu g/m^3$)		County $PM_{2.5}$ ($\mu g/m^3$)
					Mean	SD	Mean
Santa Barbara	Lompoc	054080002A07	03-29-81 to 12-31-82	91	10.631	4.649	10.627
			01-06-83 to 03-01-83	8	10.586	4.062	
Contra Costa	Richmond	056300003A07	09-24-79 to 10-14-82	167	13.920	11.333	13.920
Alameda	Livermore (Railroad Ave)	054020002A07	09-24-79 to 12-17-80	58	17.398	15.547	14.389
	Livermore (Old First St)	054020003A07	03-29-81 to 08-21-82	82	12.260	8.654	
Butte	Chico	051260002A07	03-24-82 to 08-15-82	25	9.992	4.073	15.453
			01-24-83 to 12-26-83	50	18.183	10.793	
San Francisco	San Francisco East	056860003A07	11-22-79 to 09-14-82	115	16.352	12.676	16.352
Santa Clara	San Jose	056980004A07	09-30-79 to 10-14-82	174	17.788	16.487	17.788
Fresno	Fresno (East Olive)	052800005A07	09-14-82 to 09-20-82	2	10.315	0.728	18.373
	Five Points	052820002A07	09-30-79 to 10-20-82	153	18.478	17.535	
San Diego	El Cajon	052220003A07	09-25-81 to 12-31-82	66	19.873	13.929	18.919
			01-06-83 to 08-04-83	29	16.747	11.992	
Los Angeles	Azusa (Loren Ave)	050500002A07	10-18-79 to 11-07-82	92	28.765	19.200	28.224
	West Los Angeles	054180103A07	07-08-79 to 10-14-82	147	26.752	18.667	
	Pasadena	055760004A07	10-18-79 to 12-12-81	28	34.174	16.857	
Kern	Bakersfield (Chester Ave)	050520004A07	11-17-80 to 12-01-82	53	30.803	29.190	30.863
			01-18-83 to 01-30-83	3	31.920	30.178	
Riverside	Rubidoux (Mission Blvd)	056535001A07	08-19-79 to 12-31-82	127	43.337	28.669	42.012
			01-06-83 to 06-05-83	15	30.791	33.864	

EXHIBIT D

Dose-Response:
An International Journal
January-March 2017:1-12

DOI: 10.1177/1559325817693345
journals.sagepub.com/home/dos
SAGE

Fine Particulate Matter and Total Mortality in Cancer Prevention Study Cohort Reanalysis

James E. Enstrom[1]

Abstract

Background: In 1997 the US Environmental Protection Agency (EPA) established the National Ambient Air Quality Standard (NAAQS) for fine particulate matter ($PM_{2.5}$), largely because of its positive relationship to total mortality in the 1982 American Cancer Society Cancer Prevention Study (CPS II) cohort. Subsequently, EPA has used this relationship as the primary justification for many costly regulations, most recently the Clean Power Plan. An independent analysis of the CPS II data was conducted in order to test the validity of this relationship.

Methods: The original CPS II questionnaire data, including 1982 to 1988 mortality follow-up, were analyzed using Cox proportional hazards regression. Results were obtained for 292 277 participants in 85 counties with 1979-1983 EPA Inhalable Particulate Network $PM_{2.5}$ measurements, as well as for 212 370 participants in the 50 counties used in the original 1995 analysis.

Results: The 1982 to 1988 relative risk (RR) of death from all causes and 95% confidence interval adjusted for age, sex, race, education, and smoking status was 1.023 (0.997-1.049) for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ in 85 counties and 1.025 (0.990-1.061) in the 50 original counties. The fully adjusted RR was null in the western and eastern portions of the United States, including in areas with somewhat higher $PM_{2.5}$ levels, particularly 5 Ohio Valley states and California.

Conclusion: No significant relationship between $PM_{2.5}$ and total mortality in the CPS II cohort was found when the best available $PM_{2.5}$ data were used. The original 1995 analysis found a positive relationship by selective use of CPS II and $PM_{2.5}$ data. This independent analysis of underlying data raises serious doubts about the CPS II epidemiologic evidence supporting the $PM_{2.5}$ NAAQS. These findings provide strong justification for further independent analysis of the CPS II data.

Keywords
epidemiology, $PM_{2.5}$, deaths, CPS II, reanalysis

Introduction

In 1997 the US Environmental Protection Agency (EPA) established the National Ambient Air Quality Standard (NAAQS) for fine particulate matter ($PM_{2.5}$), largely because of its positive relationship to total mortality in the 1982 American Cancer Society (ACS) Cancer Prevention Study (CPS II) cohort, as published in 1995 by Pope et al.[1] The EPA uses this positive relationship to claim that $PM_{2.5}$ *causes* premature deaths. However, the validity of this finding was immediately challenged with detailed and well-reasoned criticism.[2-4] The relationship still remains contested and much of the original criticism has never been properly addressed, particularly the need for truly independent analysis of the CPS II data.

The EPA claim that $PM_{2.5}$ *causes* premature deaths is implausible because no etiologic mechanism has ever been established and because it involves the lifetime inhalation of only about 5 g of particles that are less than 2.5 μm in diameter.[5] The $PM_{2.5}$ mortality relationship has been further challenged because the small increased risk could be due to well-known epidemiological biases, such as, the ecological fallacy, inaccurate exposure measurements, and confounding variables like copollutants. In addition, there is extensive evidence of spatial and temporal variation in $PM_{2.5}$ mortality risk (MR) that does not support 1 national standard for $PM_{2.5}$.

[1] University of California, Los Angeles and Scientific Integrity Institute, Los Angeles, CA, USA

Corresponding Author:
James E. Enstrom, University of California, Los Angeles and Scientific Integrity Institute, 907 Westwood Boulevard #200, Los Angeles, CA 90024, USA.
Email: jenstrom@ucla.edu

In spite of these serious problems, EPA and the major $PM_{2.5}$ investigators continue to assert that their positive findings are sufficient proof that $PM_{2.5}$ *causes* premature deaths. Their premature death claim has been used to justify many costly EPA regulations, most recently, the Clean Power Plan.[6] Indeed, 85% of the total estimated benefits of all EPA regulations have been attributed to reductions in $PM_{2.5}$-related premature deaths. With the assumed benefits of $PM_{2.5}$ reductions playing such a major role in EPA regulatory policy, it is essential that the relationship of $PM_{2.5}$ to mortality be independently verified with transparent data and reproducible findings.

In 1998, the Health Effects Institute (HEI) in Boston was commissioned to conduct a detailed reanalysis of the original Pope 1995 findings. The July 2000 HEI Reanalysis Report (HEI 2000) included "PART I: REPLICATION AND VALIDATION" and "PART II: SENSITIVITY ANALYSES."[7] The HEI Reanalysis Team lead by Daniel Krewski successfully replicated and validated the 1995 CPS II findings, but they did not analyze the CPS II data in ways that would determine whether the original results remained robust using different sources of air pollution data. For instance, none of their models used the best available $PM_{2.5}$ measurements as of 1995.

Particularly troubling is the fact that EPA and the major $PM_{2.5}$ investigators have ignored multiple null findings on the relationship between $PM_{2.5}$ and mortality in California. These null findings include my 2005 paper,[8] 2006 clarification,[9] 2012 American Statistical Society Joint Statistical Meeting Proceedings paper,[10] and 2015 International Conference on Climate Change presentation about the Clean Power Plan and $PM_{2.5}$-related cobenefits.[6] There is now overwhelming evidence of a null $PM_{2.5}$ mortality relationship in California dating back to 2000. The problems with the $PM_{2.5}$ mortality relationship have generated substantial scientific and political concern.

During 2011 to 2013, the US House Science, Space, and Technology Committee (HSSTC) repeatedly requested that EPA provide access to the underlying CPS II data, particularly since substantial Federal funding has been used for CPS II $PM_{2.5}$ mortality research and publications. On July 22, 2013, the HSSTC made a particularly detailed request to EPA that included 49 pages of letters dating back to September 22, 2011.[11] When EPA failed to provide the requested data, the HSSTC issued an August 1, 2013 subpoena to EPA for the CPS II data.[12] The ACS refused to comply with the HSSTC subpoena, as explained in an August 19, 2013 letter to EPA by Chief Medical Officer Otis W. Brawley.[13] Then, following the subpoena, ACS has refused to work with me and 3 other highly qualified investigators regarding collaborative analysis of the CPS II data.[14] Finally, HEI has refused to conduct my proposed CPS II analyses.[15] However, my recent acquisition of an original version of the CPS II data has made possible this first truly independent analysis.

Methods

Computer files containing the original 1982 ACS CPS II deidentified questionnaire data and 6-year follow-up data on deaths from September 1, 1982 through August 31, 1988, along with detailed documentation, were obtained from a source with appropriate access to these data, as explained in the "Acknowledgments." This article presents my initial analysis of the CPS II cohort and it is subject to the limitations of data and documentation that is not as complete and current as the data and documentation possessed by ACS.

The research described below is exempt from human participants or ethics approval because it involved only statistical analysis of existing deidentified data. Human participants' approval was obtained by ACS in 1982 when each individual enrolled in CPS II. Because of the epidemiologic importance of this analysis, an effort will be made to post on my Scientific Integrity Institute website a version of the CPS II data that fully preserves the confidentiality of all of participants and that contains enough information to verify my findings.

Of the 1.2 million total CPS II participants, analysis has been done on 297 592 participants residing in 85 counties in the continental United States with 1979 to 1983 EPA Inhalable Particulate Network (IPN) $PM_{2.5}$ measurements.[16,17] Among these participants, there were 18 612 total deaths from September 1, 1982 through August 31, 1988; 17 329 of these deaths (93.1%) had a known date of death. Of the 297 592 participants, 292 277 had age at entry of 30 to 99 years and sex of male [1] or female [2]. Of the 292 277 participants, 269 766 had race of white [1,2,5] or black [3,4]; education level of no or some high school [1,2], high school graduate [3], some college [4,5], college graduate [6], or graduate school [7]; and smoking status of never [1], former [5-8 for males and 3 for females], or current [2-4 for males and 2 for females]. Those participants reported to be dead [D, G, K] but without an exact date of death have been assumed to be alive in this analysis. The unconfirmed deaths were randomly distributed and did not impact relative comparisons of death in a systematic way. The computer codes for the above variables are shown in brackets.

CPS II participants were entered into the master data file geographically. Since this deidentified data file does not contain home addresses, the Division number and Unit number assigned by ACS to each CPS II participant have been used to define their county of residence. For instance, ACS Division 39 represents the state of Ohio and its Unit 041 represents Jefferson County, which includes the city of Steubenville, where the IPN $PM_{2.5}$ measurements were made. In other words, most of the 575 participants in Unit 041 lived in Jefferson County as of September 1, 1982. The IPN $PM_{2.5}$ value of 29.6739 $\mu g/m^3$, based on measurements made in Steubenville, was assigned to all CPS II participants in Unit 041. This $PM_{2.5}$ value is a weighted average of 53 measurements (mean of 33.9260 $\mu g/m^3$) and 31 measurements (mean of 29.4884 $\mu g/m^3$) made during 1979 to 1982[16] and 53 measurements (mean of 27.2473 $\mu g/m^3$) and 54 measurements (mean of 28.0676 $\mu g/m^3$) made during 1983.[17] The IPN $PM_{2.5}$ data were collected only during 1979 to 1983, although some other IPN air pollution data were collected through 1984. The values for each county that includes a city with CPS II participants and IPN $PM_{2.5}$ measurements are shown in Appendix Table A1.

Table 1. Summary Characteristics of CPS II Participants in (1) Pope 1995 Table 1,[1] (2) HEI 2000 Table 24,[7] and (3) Current Analysis Based on CPS II Participants in 50 and 85 Counties.

Characteristics	Pope 1995 Table 1	HEI 2000 Table 24	Current CPS II Analysis		
			n = 50 HEI $PM_{2.5}$	n = 50 IPN $PM_{2.5}$	n = 85 IPN $PM_{2.5}$
Number of metro areas	50	50			
Number of counties	Not stated	Not stated	50	50	85
Age–sex-adjusted participants			212 370	212 370	292 277
Fully adjusted participants	295 223	298 817	195 215	195 215	269 766
Age–sex-adjusted deaths			12 518	12 518	17 231
Fully adjusted deaths	20 765	23 093	11 221	11 221	15 593
Values below are for participants in fully adjusted results					
Age at enrollment, mean years	56.6	56.6	56.66	56.66	56.64
Sex (% females)	55.9	56.4	56.72	56.72	56.61
Race (% white)	94.0	94.0	94.58	94.58	95.09
Less than high school education, %	11.3	11.3	11.71	11.71	11.71
Never smoked regularly, %			41.69	41.69	41.57
Former smoker, %			33.25	33.25	33.67
Former cigarette smoker, %	29.4	30.2	30.43	30.43	30.81
Current smoker, %			25.06	25.06	24.76
Current cigarette smoker, %	21.6	21.4	21.01	21.01	20.76
Fine particles, $\mu g/m^3$					
Average	18.2	18.2	17.99	21.37	21.16
SD	5.1	4.4	4.52	5.30	5.98
Range	9.0-33.5	9.0-33.4	9.0-33.4	10.77-29.67	10.63-42.01

Abbreviations: CPS, Cancer Prevention Study; HEI, Health Effects Institute; IPN, Inhalable Particulate Network; $PM_{2.5}$, fine particulate matter.

To make the best possible comparison with Pope 1995 and HEI 2000 results, the HEI $PM_{2.5}$ value of 23.1 $\mu g/m^3$ for Steubenville was assigned to all participants in Unit 041. This value is the median of $PM_{2.5}$ measurements made in Steubenville and is shown in HEI 2000 Appendix D "Alternative Air Pollution Data in the ACS Study."[7] Analyses were done for the 50 counties containing the original 50 cities with CPS II participants and HEI $PM_{2.5}$ values used in Pope 1995 and HEI 2000. Additional analyses were done for all 85 counties containing cities with both CPS II participants and IPN $PM_{2.5}$ data. Without explanation, Pope 1995 and HEI 2000 omitted from their analyses, 35 cities with CPS II participants and IPN $PM_{2.5}$ data. To be clear, these analyses are based on the CPS II participants assigned to each Unit (county) that included a city with IPN $PM_{2.5}$ data. The original Pope 1995 and HEI 2000 analyses were based on the CPS II participants assigned to each metropolitan area (MA) that included a city with HEI $PM_{2.5}$ data, as defined in HEI 2000 Appendix F "Definition of Metropolitan Areas in the ACS Study."[7] The MA, which was equivalent to the US Census Bureau Standard Metropolitan Statistical Area (SMSA), always included the county containing the city with the HEI $PM_{2.5}$ data and often included 1 or more additional counties.

The SAS 9.4 procedure PHREG was used to conduct Cox proportional hazards regression.[18] Relative risks (RRs) for death from all causes and 95% confidence intervals (CI) were calculated using age–sex adjustment and full adjustment (age, sex, race, education, and smoking status, as defined above). Each of the 5 adjustment variables had a strong relationship to total mortality. Race, education, and smoking status were the 3 adjustment variables that had the greatest impact on the age–sex-adjusted RR. The Pope 1995 and HEI 2000 analyses used 4 additional adjustment variables that had a lesser impact on the age–sex-adjusted RR.

In addition, county-level ecological analyses were done by comparing IPN $PM_{2.5}$ and HEI $PM_{2.5}$ values to 1980 age-adjusted white total death rates (DRs) determined by the Centers for Disease Control and Prevention (CDC) WONDER[19] and mortality risks (MRs) as shown in Figures 5 and 21 of HEI 2000.[7] Death rates are age adjusted to the 2000 US Standard Population and are expressed as annual deaths per 100 000 persons. The SAS 9.4 procedure REGRESSION was used to conduct linear regression of $PM_{2.5}$ values with DRs and MRs.

Appendix Table A1 lists the 50 original cities used in Pope 1995 and HEI 2000 and includes city, county, state, ACS Division and Unit numbers, Federal Information Processing Standards (FIPS) code, IPN average $PM_{2.5}$ level, HEI median $PM_{2.5}$ level, 1980 DR, and HEI MR. Appendix Table A1 also lists similar information for the 35 additional cities with CPS II participants and IPN $PM_{2.5}$ data. However, HEI $PM_{2.5}$ and HEI MR data are not available for these 35 cities.

Results

Table 1 shows basic demographic characteristics for the CPS II participants, as stated in Pope 1995,[1] HEI 2000,[7] and this current analysis. There is excellent agreement on age, sex, race, education, and smoking status. However, the IPN $PM_{2.5}$ averages are generally about 20% higher than the HEI $PM_{2.5}$ medians, although the differences range from +78% to −28%.

Table 2. Age–Sex-Adjusted and Fully Adjusted Relative Risk of Death From All Causes (RR and 95% CI) From September 1, 1982 Through August 31, 1988 Associated With Change of 10 $\mu g/m^3$ Increase in $PM_{2.5}$ for CPS II Participants Residing in 50 and 85 Counties in the Continental United States With 1979 to 1983 IPN $PM_{2.5}$ Measurements.[a]

$PM_{2.5}$ Years and Source	Number of Counties	Number of Participants	Number of Deaths	RR	95% CI Lower Upper	Average $PM_{2.5}$
Age–sex adjusted RR for the continental United States						
1979-1983 IPN	85	292 277	17 321	1.038	(1.014-1.063)	21.16
1979-1983 IPN	50	212 370	12 518	1.046	(1.013-1.081)	21.36
1979-1983 HEI	50	212 370	12 518	1.121	(1.078-1.166)	17.99
Fully adjusted RR for the continental United States						
1979-1983 IPN	85	269 766	15 593	1.023	(0.997-1.049)	21.15
1979-1983 IPN	50	195 215	11 221	1.025	(0.990-1.061)	21.36
1979-1983 HEI	50	195 215	11 221	1.082	(1.039-1.128)	17.99
Age–sex adjusted RR for Ohio Valley States (IN, KY, OH, PA, WV)						
1979-1983 IPN	17	56 979	3649	1.126	(1.011-1.255)	25.51
1979-1983 IPN	12	45 303	2942	1.079	(0.951-1.225)	25.76
1979-1983 HEI	12	45 303	2942	1.153	(1.027-1.296)	22.02
Fully adjusted RR for Ohio Valley states (IN, KY, OH, PA, WV)						
1979-1983 IPN	17	53 026	3293	1.096	(0.978-1.228)	25.51
1979-1983 IPN	12	42 174	2652	1.050	(0.918-1.201)	25.75
1979-1983 HEI	12	42 174	2652	1.111	(0.983-1.256)	22.02
Age–sex adjusted RR for states other than the Ohio Valley states						
1979-1983 IPN	68	235 298	13 672	0.999	(0.973-1.027)	20.11
1979-1983 IPN	38	167 067	9576	0.983	(0.946-1.021)	20.18
1979-1983 HEI	38	167 067	9576	1.045	(0.997-1.096)	16.90
Fully adjusted RR for states other than the Ohio Valley states						
1979-1983 IPN	68	216 740	12 300	0.994	(0.967-1.023)	20.09
1979-1983 IPN	38	153 041	8569	0.975	(0.936-1.015)	20.15
1979-1983 HEI	38	153 041	8569	1.025	(0.975-1.078)	16.89

Abbreviations: CI, confidence interval; CPS, Cancer Prevention Study; HEI, Health Effects Institute; IPN, Inhalable Particulate Network; $PM_{2.5}$, particulate matter.
[a]Analysis includes continental United States, 5 Ohio Valley states, and remainder of the states. Appendix Table A1 lists the 85 cities and counties with $PM_{2.5}$ measurements.

Table 2 shows that during 1982 to 1988, there was no significant relationship between IPN $PM_{2.5}$ and total mortality in the entire United States. The fully adjusted RR and 95% CI was 1.023 (0.997-1.049) for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ in all 85 counties and 1.025 (0.990-1.061) in the 50 original counties. Indeed, the fully adjusted RR was not significant in any area of the United States, such as, the states west of the Mississippi River, the states east of the Mississippi River, the 5 Ohio Valley states (Indiana, Kentucky, Ohio, Pennsylvania, and West Virginia), and the states other than the Ohio Valley states. The age–sex-adjusted and fully adjusted RRs in the states other than the Ohio Valley states are all consistent with no relationship and most are very close to 1.00. The slightly positive age–sex-adjusted RRs for the entire United States and the Ohio Valley states became statistically consistent with no relationship after controlling for the 3 confounding variables of race, education, and smoking status.

However, the fully adjusted RR for the entire United States was 1.082 (1.039-1.128) when based on the HEI $PM_{2.5}$ values in 50 counties. This RR agrees quite well with the fully adjusted RR of 1.067 (1.037-1.099) for 1982 to 1989, which is shown in Table 34 of the June 2009 HEI Extended Follow-up Research Report (HEI 2009).[20] Thus, the positive nationwide RRs in the CPS II cohort depend upon the use of HEI $PM_{2.5}$ values. The nationwide RRs are consistent with no effect when based on IPN $PM_{2.5}$ values. The findings in Table 2 clearly demonstrate the large influence of $PM_{2.5}$ values and geography on the RRs.

Table 3 shows that the fully adjusted RR in California was 0.992 (0.954-1.032) when based on IPN $PM_{2.5}$ values in all 11 California counties. This null finding is consistent with the 15 other findings of a null relationship in California, all of which are shown in Appendix Table B1. However, when the RR is based on the 4 California counties used in Pope 1995 and HEI 2000, there is a significant inverse relationship. The fully adjusted RR is 0.879 (0.805-0.960) when based on the IPN $PM_{2.5}$ values and is 0.870 (0.788-0.960) when based on the HEI $PM_{2.5}$ values. This significant inverse relationship is in exact agreement with the finding of a special analysis of the CPS II cohort done for HEI by Krewski in 2010, which yielded a fully adjusted RR of 0.872 (0.805-0.944) during 1982 to 1989 in California when based on HEI $PM_{2.5}$ values.[21] In this instance, the California RRs are clearly dependent upon the number of counties used.

Table 4 shows that the ecological analysis based on linear regression is quite consistent with the proportional hazard regression results in Tables 2 and 3, in spite of the fact that the regression results are not fully adjusted. Using 1980 age-adjusted white total DRs versus HEI $PM_{2.5}$ values in 50 counties, linear regression yielded a regression coefficient of 6.96 (standard error [SE] = 1.85) that was statistically significant at the 95% confidence level. Pope 1995 reported a significant regression coefficient for 50 cities of 8.0 (SE = 1.4). However, this positive coefficient is

Table 3. Age–Sex-Adjusted and Fully Adjusted Relative Risk of Death From All Causes (RR and 95% CI) From September 1, 1982 Through August 31, 1988 Associated With 10 $\mu g/m^3$ Increase in $PM_{2.5}$ for California CPS II Participants Living in 4 and 11 Counties With 1979 to 1983 IPN $PM_{2.5}$ Measurements.[a]

$PM_{2.5}$ Years and Source	Number of Counties	Number of Participants	Number of Deaths	RR	95% CI of RR Lower Upper	Average $PM_{2.5}$
Age–sex adjusted RR for California during 1982 to 1988						
1979-1983 IPN	11	66 615	3856	1.005	(0.968-1.043)	24.08
1979-1983 IPN	4	40 527	2146	0.904	(0.831-0.983)	24.90
1979-1983 HEI	4	40 527	2146	0.894	(0.817-0.986)	18.83
Fully adjusted (age, sex, race, education, and smoking status) RR for California during 1982 to 1988						
1979-1983 IPN	11	60 521	3512	0.992	(0.954-1.032)	24.11
1979-1983 IPN	4	36 201	1939	0.879	(0.805-0.960)	25.01
1979-1983 HEI	4	36 201	1939	0.870	(0.788-0.960)	18.91
Fully adjusted (44 confounders) RR for California during 1982 to 1989 as per Krewski[21]						
"Same" Standard Cox Model 1979-1983 HEI	4	40 408		0.872	(0.805-0.944)	~19
"Different" Standard Cox Model 1979-1983 HEI	4	38 925		0.893	(0.823-0.969)	~19

Abbreviations: CI, confidence interval; CPS, Cancer Prevention Study; HEI, Health Effects Institute; IPN, Inhalable Particulate Network; $PM_{2.5}$, particulate matter.
[a]Also, fully adjusted RR for California participants in 4 counties from September 1, 1982 through December 31, 1989 as calculated by Krewski.[21]

Table 4. Linear Regression Results for 1979 to 1983 IPN $PM_{2.5}$ and 1979 to 1983 HEI $PM_{2.5}$ Versus 1980 Age-Adjusted White Total Death Rate (DR) for 85 Counties With IPN $PM_{2.5}$ Data and for 50 HEI 2000 Counties With IPN $PM_{2.5}$ and HEI $PM_{2.5}$ data.

DR or MR, $PM_{2.5}$ Years and Source	Number of Counties	DR or MR Intercept	DR or MR Slope	95% CI of DR or MR Slope Lower	Upper	*P* Value
Entire continental United States						
DR and 1979-1983 IPN	85	892.68	6.8331	3.8483	9.8180	0.0000
DR and 1979-1983 HEI	50	910.92	6.9557	3.2452	10.6662	0.0004
MR and 1979-1983 IPN	50	0.6821	0.0102	0.0044	0.0160	0.0009
MR and 1979-1983 HEI	50	0.6754	0.0121	0.0068	0.0173	0.0000
Ohio Valley states (IN, KY, OH, PA, and WV)						
DR and 1979-1983 IPN	17	941.77	6.0705	−0.0730	12.2139	0.0524
DR and 1979-1983 HEI	12	1067.29	1.3235	−7.3460	9.9930	0.7408
MR and 1979-1983 IPN	12	0.8153	0.0077	−0.0054	0.0208	0.2202
MR and 1979-1983 HEI	12	0.9628	0.0020	−0.0080	0.0121	0.6608
States other than the Ohio Valley states						
DR and 1979-1983 IPN	68	921.45	4.8639	0.9093	8.8186	0.0167
DR and 1979-1983 HEI	38	934.66	4.8940	−0.4337	10.2218	0.0706
MR and 1979-1983 IPN	38	0.8111	0.0020	−0.0054	0.0094	0.5891
MR and 1979-1983 HEI	38	0.7334	0.0072	0.0000	0.0144	0.0491
States west of the Mississippi river						
DR and 1979-1983 IPN	36	920.10	4.0155	−0.9396	8.9706	0.1088
DR and 1979-1983 HEI	22	930.11	4.1726	−5.2015	13.5468	0.3642
MR and 1979-1983 IPN	22	0.8663	−0.0025	−0.0162	0.0112	0.7067
MR and 1979-1983 HEI	22	0.6413	0.0134	−0.0018	0.0285	0.0807
California						
DR and 1979-1983 IPN	11	921.71	3.6516	−1.8230	9.1262	0.1656
DR and 1979-1983 HEI	4	992.50	1.9664	−46.6929	50.6256	0.8780
MR and 1979-1983 IPN	4	0.9529	−0.0074	−0.0600	0.0453	0.6072
MR and 1979-1983 HEI	4	0.8336	−0.0021	−0.0618	0.0576	0.8935

Abbreviations: CI, confidence interval; HEI, Health Effects Institute; IPN, Inhalable Particulate Network; MR, mortality risk; $PM_{2.5}$, particulate matter.
[a]Linear regression results are also shown for 1979 to 1983 IPN $PM_{2.5}$ and 1979 to 1983 HEI $PM_{2.5}$ versus MR for the 50 "cities" (metropolitan areas) in figures 5 and 21 in HEI 2000.

misleading because both DRs and $PM_{2.5}$ levels are higher in the East than in the West. Regional regression analyses did not generally yield significant regression coefficients. Specifically, there were no significant regression coefficients for California, the 5 Ohio Valley states, or all states west of the Mississippi River. These findings reinforce the CPS II cohort evidence of statistically insignificant $PM_{2.5}$ MR throughout the United States.

Conclusion

This independent analysis of the CPS II cohort found that there was no significant relationship between $PM_{2.5}$ and death from all causes during 1982 to 1988, when the best available $PM_{2.5}$ measurements were used for the 50 original counties and for all 85 counties with $PM_{2.5}$ data and CPS II participants. However, a positive relationship was found when the HEI $PM_{2.5}$ measurements were used for the 50 original counties, consistent with the findings in Pope 1995 and HEI 2000. This null and positive evidence demonstrates that the $PM_{2.5}$ mortality relationship is not robust and is quite sensitive to the $PM_{2.5}$ data and CPS II participants used in the analysis.

Furthermore, the following statement on page 80 of HEI 2000 raises serious doubts about the quality of the air pollution data used in Pope 1995 and HEI 2000: "AUDIT OF AIR QUALITY DATA. The ACS study was not originally designed as an air pollution study. The air quality monitoring data used for the ACS analyses came from various sources, some of which are now technologically difficult to access. Documentation of the statistical reduction procedures has been lost. Summary statistics for different groups of standard metropolitan statistical areas had been derived by different investigators. These data sources do not indicate whether the tabulated values refer to all or a subset of monitors in a region or whether they represent means or medians."[7]

The Pope 1995 and HEI 2000 analyses were based on 50 median $PM_{2.5}$ values shown in Appendix A of the 1988 Brookhaven National Laboratory Report 52122 by Lipfert et al.[22] These analyses did not use or cite the high quality and widely known EPA IPN $PM_{2.5}$ data in spite of the fact that these data have been available in 2 detailed EPA reports since 1986.[16,17] Lipfert informed HEI about the IPN data in 1998: "During the early stages of the Reanalysis Project, I notified HEI and the reanalysis contractors of the availability of an updated version of the IPN data from EPA, which they apparently obtained. This version includes more locations and a slightly longer period of time. It does not appear that the newer IPN data are listed in Appendix G, and it is thus not possible to confirm if SMSA assignments were made properly."[23]

Thus, the HEI Reanalysis Team failed to properly "evaluate the sensitivity of the original findings to the indicators of exposure to fine particle air pollution used by the Original Investigators" and failed to select "all participants who lived within each MA for which data on sulfate or fine particle pollution were available."[7] Furthermore, HEI 2009 did not use these data even though the investigators were aware of my 2005 null $PM_{2.5}$ mortality findings in California,[8] which were based on the IPN data for 11 California counties, instead of the 4 California counties used in Pope 1995 and HEI 2000. Indeed, HEI 2009 did not cite my 2005 findings, in spite of my personal discussion of these findings with Pope, Jerrett, and Burnett on July 11, 2008.[24] Finally, HEI 2009 did not acknowledge or address my 2006 concerns about the geographic variation in $PM_{2.5}$ MR clearly shown in HEI 2000 Figure 21,[7] which is included here as Appendix Figure C1. HEI 2009 entirely avoided the issue of geographic variation in $PM_{2.5}$ MR and omitted the equivalent to HEI 2000 Figure 21.

Since 2002, HEI has repeatedly refused to provide the city-specific $PM_{2.5}$-related MR for the 50 cities included in HEI 2000 Figure 21.[15] I estimated these MRs in 2010 based on visual measurements of HEI 2000 Figure 5, and my estimates are shown in Appendix Table A1.[25] Figure 21 and its MRs represented early evidence that there was no $PM_{2.5}$-related MR in California. Appendix Table B1 shows the now overwhelming 2000 to 2016 evidence from 6 different cohorts that there is no relationship between $PM_{2.5}$ and total mortality in California. Indeed, the weighted average RR of the latest results from the 6 California cohorts is RR = 0.999 (0.988-1.010).[26]

The authors of the CPS II $PM_{2.5}$ mortality publications, which began with Pope 1995, have faced original criticism,[2-4] my criticism,[6-10,14,15] and the criticism of the HSSTC and its subpoena.[11-13] Now, my null findings represent a direct challenge to the positive findings of Pope 1995. All of this criticism is relevant to the EPA claim that $PM_{2.5}$ has a *causal* relationship to total mortality. The authors of Pope 1995, HEI 2000, and HEI 2009 need to promptly address my findings, as well as the earlier criticism. Then, they need to cooperate with critics on transparent air pollution epidemiology analyses of the CPS II cohort data.

Also, major scientific journals like the *New England Journal of Medicine* (*NEJM*) and *Science*, which have consistently written about the positive relationship between $PM_{2.5}$ and total mortality, need to publish evidence of no relationship when strong null evidence is submitted to them. In 2015, *Science* immediately rejected without peer reviewing 3 versions of strong evidence that $PM_{2.5}$ does not *cause* premature deaths.[5] In 2016, *Science* immediately rejected without peer reviewing this article. Indeed, this article was rejected by *NEJM*, *Science*, and 5 other major journals, as described in a detailed compilation of relevant correspondence.[27] Most troubling is the rejection by the *American Journal of Respiratory and Clinical Care Medicine*, which has published Pope 1995 and several other $PM_{2.5}$ mortality articles based on the CPS II cohort data.

In summary, the null CPS II $PM_{2.5}$ mortality findings in this article directly challenge the original positive Pope 1995 findings, and they raise serious doubts about the CPS II epidemiologic evidence supporting the $PM_{2.5}$ NAAQS. These findings demonstrate the importance of independent and transparent analysis of underlying data. Finally, these findings provide strong justification for further independent analysis of CPS II cohort data.

Appendix A

Table A1. List of the 85 Counties Containing the 50 Cities Used in Pope 1995, HEI 2000, and This Analysis, as well as the 35 Additional Cities Used Only in This Analysis.[a]

State	ACS Div-Unit	FIPS Code	IPN/HEI County Containing IPN/HEI City	IPN/HEI City With $PM_{2.5}$ Measurements	1979-1983 IPN $PM_{2.5}$, $\mu g/m^3$, (Weighted Average)	1979-1983 HEI $PM_{2.5}$, $\mu g/m^3$ (Median)	1980 Age-Adj White Death Rate (DR)	HEI Figure 5 Mortality Risk (MR)
AL	01037	01073	Jefferson	Birmingham	25.6016	24.5	1025.3	0.760
AL	01049	01097	Mobile	Mobile	22.0296	20.9	1067.2	0.950
AZ	03700	04013	Maricopa	Phoenix	15.7790	15.2	953.0	0.855
AR	04071	05119	Pulaski	Little Rock	20.5773	17.8	1059.4	0.870
CA	06001	06001	Alameda	Livermore	14.3882		1016.6	
CA	06002	06007	Butte	Chico	15.4525		962.5	
CA	06003	06013	Contra Costa	Richmond	13.9197		937.1	
CA	06004	06019	Fresno	Fresno	18.3731	10.3	1001.4	0.680
CA	06008	06029	Kern	Bakersfield	30.8628		1119.3	
CA	06051	06037	Los Angeles	Los Angeles	28.2239	21.8	1035.1	0.760
CA	06019	06065	Riverside	Rubidoux	42.0117		1013.9	
CA	06020	06073	San Diego	San Diego	18.9189		943.7	
CA	06021	06075	San Francisco	San Francisco	16.3522	12.2	1123.1	0.890
CA	06025	06083	Santa Barbara	Lompoc	10.6277		892.8	
CA	06026	06085	Santa Clara	San Jose	17.7884	12.4	921.9	0.885
CO	07004	08031	Denver	Denver	10.7675	16.1	967.3	0.925
CO	07047	08069	Larimer	Fort Collins	11.1226		810.5	
CO	07008	08101	Pueblo	Pueblo	10.9155		1024.1	
CT	08001	09003	Hartford	Hartford	18.3949	14.8	952.0	0.845
CT	08004	09005	Litchfield	Litchfield	11.6502		941.5	
DE	09002	10001	Kent	Dover	19.5280		959.4	
DE	09004	10003	New Castle	Wilmington	20.3743		1053.7	
DC	10001	11001	Dist Columbia	Washington	25.9289	22.5	993.2	0.850
FL	11044	12057	Hillsborough	Tampa	13.7337	11.4	1021.8	0.845
GA	12027	13051	Chatham	Savannah	17.8127		1029.6	
GA	12062	13121	Fulton	Atlanta	22.5688	20.3	1063.5	0.840
ID	13001	16001	ADA	Boise	18.0052	12.1	892.6	0.600
IL	14089	17031	Cook	Chicago	25.1019	21.0	1076.3	0.945
IL	14098	17197	Will	Braidwood	17.1851		1054.0	
IN	15045	18089	Lake	Gary	27.4759	25.2	1129.8	0.995
IN	15049	18097	Marion	Indianapolis	23.0925	21.1	1041.2	0.970
KS	17287	20173	Sedgwick	Wichita	15.0222	13.6	953.4	0.890
KS	17289	20177	Shawnee	Topeka	11.7518	10.3	933.7	0.830
KY	18010	21019	Boyd	Ashland	37.7700		1184.6	
KY	18055	21111	Jefferson	Louisville	24.2134		1095.7	
MD	21106	24510	Baltimore City	Baltimore	21.6922		1237.8	
MD	21101	24031	Montgomery	Rockville	20.2009		881.9	
MA	22105	25013	Hampden	Springfield	17.5682		1025.3	
MA	22136	25027	Worcester	Worcester	16.2641		1014.6	
MN	25001	27053	Hennepin	Minneapolis	15.5172	13.7	905.3	0.815
MN	25150	27123	Ramsey	St Paul	15.5823		935.7	
MS	26086	28049	Hinds	Jackson	18.1339	15.7	1087.4	0.930
MO	27001	29095	Jackson	Kansas City	17.8488		1090.3	
MT	28009	30063	Missoula	Missoula	17.6212		938.0	
MT	28011	30093	Silver Bow	Butte	16.0405		1299.5	
NE	30028	31055	Douglas	Omaha	15.2760	13.1	991.0	0.880
NV	31101	32031	Washoe	Reno	13.1184	11.8	1049.5	0.670
NJ	33004	34007	Camden	Camden	20.9523		1146.9	
NJ	33007	34013	Essex	Livingston	16.4775		1072.7	
NJ	33009	34017	Hudson	Jersey City	19.9121	17.3	1172.6	0.810
NM	34201	35001	Bernalillo	Albuquerque	12.8865	9.0	1014.7	0.710
NY	36014	36029	Erie	Buffalo	25.1623	23.5	1085.6	0.960
NY	35001	36061	New York	New York City	23.9064		1090.4	
NC	37033	37063	Durham	Durham	19.4092	16.8	1039.2	1.000

(continued)

Table A1. (continued)

State	ACS Div-Unit	FIPS Code	IPN/HEI County Containing IPN/HEI City	IPN/HEI City With $PM_{2.5}$ Measurements	1979-1983 IPN $PM_{2.5}$, $\mu g/m^3$, (Weighted Average)	1979-1983 HEI $PM_{2.5}$, $\mu g/m^3$ (Median)	1980 Age-Adj White Death Rate (DR)	HEI Figure 5 Mortality Risk (MR)
NC	37064	37119	Mecklenburg	Charlotte	24.1214	22.6	932.8	0.835
OH	39009	39017	Butler	Middletown	25.1789		1108.3	
OH	39018	39035	Cuyahoga	Cleveland	28.4120	24.6	1089.1	0.980
OH	39031	39061	Hamilton	Cincinnati	24.9979	23.1	1095.2	0.980
OH	39041	39081	Jefferson	Steubenville	29.6739	23.1	1058.6	1.145
OH	39050	39099	Mahoning	Youngstown	22.9404	20.2	1058.4	1.060
OH	39057	39113	Montgomery	Dayton	20.8120	18.8	1039.5	0.980
OH	39077	39153	Summit	Akron	25.9864	24.6	1064.0	1.060
OK	40055	40109	Oklahoma	Oklahoma City	14.9767	15.9	1050.4	0.985
OR	41019	41039	Lane	Eugene	17.1653		885.5	
OR	41026	41051	Multnomah	Portland	16.3537	14.7	1060.8	0.830
PA	42101	42003	Allegheny	Pittsburgh	29.1043	17.9	1115.6	1.005
PA	42443	42095	Northampton	Bethlehem	19.5265		998.6	
PA	43002	42101	Philadelphia	Philadelphia	24.0704	21.4	1211.0	0.910
RI	45001	44007	Providence	Providence	14.2341	12.9	1006.1	0.890
SC	46016	45019	Charleston	Charleston	16.1635		1023.5	
TN	51019	47037	Davidson	Nashville	21.8944	20.5	981.9	0.845
TN	51088	47065	Hamilton	Chattanooga	18.2433	16.6	1087.9	0.840
TX	52811	48113	Dallas	Dallas	18.7594	16.5	1024.9	0.850
TX	52859	48141	El Paso	El Paso	16.9021	15.7	903.5	0.910
TX	52882	48201	Harris	Houston	18.0421	13.4	1025.7	0.700
UT	53024	49035	Salt Lake	Salt Lake City	16.6590	15.4	954.3	1.025
VA	55024	51059	Fairfax	Fairfax	19.5425		925.7	
VA	55002	51710	Norfolk City	Norfolk	19.5500	16.9	1139.3	0.910
WA	56017	53033	King	Seattle	14.9121	11.9	943.6	0.780
WA	56032	53063	Spokane	Spokane	13.5200	9.4	959.2	0.810
WV	58130	54029	Hancock	Weirton	25.9181		1094.8	
WV	58207	54039	Kanawha	Charleston	21.9511	20.1	1149.5	1.005
WV	58117	54069	Ohio	Wheeling	23.9840	33.4	1117.5	1.020
WI	59005	55009	Brown	Green Bay	20.5462		931.0	
WI	59052	55105	Rock	Beloit	19.8584		1019.4	

[a]Each location includes State, ACS Division Unit number, Federal Information Processing Standards (FIPS) code, IPN/HEI county, IPN/HEI city with $PM_{2.5}$ measurements, 1979-1983 IPN average $PM_{2.5}$ level, 1979-1983 HEI median $PM_{2.5}$ level, 1980 age-adjusted white county total death rate (annual deaths per 100 000), and HEI 2000 figure 5 mortality risk for HEI city (metropolitan area). List also includes 35 additional counties containing cities with IPN $PM_{2.5}$ data used in this analysis. These 35 counties do not have HEI $PM_{2.5}$ data.

Appendix B

Table B1. Epidemiologic Cohort Studies of $PM_{2.5}$ and Total Mortality in California, 2000 to 2016: Relative Risk of Death From All Causes (RR and 95% CI) Associated With Increase of 10 $\mu g/m^3$ in $PM_{2.5}$ (http://scientificintegrityinstitute.org/NoPMDeaths081516.pdf).

Krewski 2000 and 2010[a,b]	CA CPS II Cohort	N = 40 408	RR = 0.872 (0.805-0.944)	1982-1989
(N = [18 000 M + 22 408 F]; 4 MSAs; 1979-1983 $PM_{2.5}$; 44 covariates)				
McDonnell 2000[c]	CA AHSMOG Cohort	N ~ 3800	RR ~ 1.00 (0.95-1.05)	1977-1992
(N ~ [1347 M + 2422 F]; SC&SD&SF AB; M RR = 1.09 (0.98-1.21) & F RR ~ 0.98 (0.92-1.03))				
Jerrett 2005[d]	CPS II Cohort in LA Basin	N = 22 905	RR = 1.11 (0.99-1.25)	1982-2000
(N = 22 905 M and F; 267 zip code areas; 1999-2000 $PM_{2.5}$; 44 cov + max confounders)				
Enstrom 2005[e]	CA CPS I Cohort	N = 35 783	RR = 1.039 (1.010-1.069)	1973-1982
(N = [15 573 M + 20 210 F]; 11 counties; 1979-1983 $PM_{2.5}$)			RR = 0.997 (0.978-1.016)	1983-2002
Enstrom 2006[f]	CA CPS I Cohort	N = 35 783	RR = 1.061 (1.017-1.106)	1973-1982
(N = [15 573 M + 20 210 F]; 11 counties; 1979-1983 and 1999-2001 $PM_{2.5}$)			RR = 0.995 (0.968-1.024)	1983-2002
Zeger 2008[g]	MCAPS Cohort "West"	N = 3 100 000	RR = 0.989 (0.970-1.008)	2000-2005
(N = [1.5 M M + 1.6 M F]; Medicare enrollees in CA + OR + WA (CA = 73%); 2000-2005 $PM_{2.5}$)				

(continued)

Table B1. (continued)

Jerrett 2010[h]	CA CPS II Cohort	N = 77 767	RR ~ 0.994 (0.965-1.025)	1982-2000
(N = [34 367 M + 43 400 F]; 54 counties; 2000 $PM_{2.5}$; KRG ZIP; 20 ind cov + 7 eco var; slide 12)				
Krewski 2010[b] (2009)	CA CPS II Cohort			
(4 MSAs; 1979-1983 $PM_{2.5}$; 44 cov)		N = 40 408	RR = 0.960 (0.920-1.002)	1982-2000
(7 MSAs; 1999-2000 $PM_{2.5}$; 44 cov)		N = 50 930	RR = 0.968 (0.916-1.022)	1982-2000
Jerrett 2011[i]	CA CPS II Cohort	N = 73 609	RR = 0.994 (0.965-1.024)	1982-2000
(N = [32 509 M + 41 100 F]; 54 counties; 2000 $PM_{2.5}$; KRG ZIP Model; 20 ind cov + 7 eco var; Table 28)				
Jerrett 2011[i]	CA CPS II Cohort	N = 73 609	RR = 1.002 (0.992-1.012)	1982-2000
(N = [32 509 M + 41 100 F]; 54 counties; 2000 $PM_{2.5}$; Nine Model Ave; 20 ic + 7 ev; Figure 22 and Tables 27-32)				
Lipsett 2011[j]	CA Teachers Cohort	N = 73 489	RR = 1.01 (0.95-1.09)	2000-2005
(N = [73 489 F]; 2000-2005 $PM_{2.5}$)				
Ostro 2011[k]	CA Teachers Cohort	N = 43 220	RR = 1.06 (0.96-1.16)	2002-2007
(N = [43 220 F]; 2002-2007 $PM_{2.5}$)				
Jerrett 2013[l]	CA CPS II Cohort	N = 73 711	RR = 1.060 (1.003-1.120)	1982-2000
(N = [~32 550 M + ~41 161 F]; 54 counties; 2000 $PM_{2.5}$; LUR Conurb Model; 42 ind cov + 7 eco var + 5 metro; Table 6)				
Jerrett 2013[l]	CA CPS II Cohort	N = 73 711	RR = 1.028 (0.957-1.104)	1982-2000
(Same parameters and model as above, except including co-pollutants NO_2 and Ozone; Table 5)				
Ostro 2015[m]	CA Teachers Cohort	N = 101 884	RR = 1.01 (0.98-1.05)	2001-2007
(N = [101 881 F]; 2002-2007 $PM_{2.5}$) (all natural causes of death)				
Thurston 2016[n]	CA NIH-AARP Cohort	N = 160 209	RR = 1.02 (0.99-1.04)	2000-2009
(N = [~95 965 M + ~64 245 F]; full baseline model: $PM_{2.5}$ by zip code; Table 3) (all natural causes of death)				
Enstrom 2016 unpublished	CA NIH-AARP Cohort	N = 160 368	RR = 1.001 (0.949-1.055)	2000-2009
(N = [~96 059 M + ~64 309 F]; full baseline model: 2000 $PM_{2.5}$ by county)				

[a]Krewski D. "Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: HEI Special Report. July 2000". 2000. Figure 5 and Figure 21 of Part II: Sensitivity Analyses http://www.scientificintegrityinstitute.org/HEIFigure5093010.pdf.

[b]Krewski D. August 31, 2010 letter from Krewski to Health Effects Institute and CARB with California-specific $PM_{2.5}$ mortality results from Table 34 in Krewski 2009. 2010. http://www.arb.ca.gov/research/health/pm-mort/HEI_Correspondence.pdf

[c]McDonnell WF, Nishino-Ishikawa N, Petersen FF, Chen LH, Abbey DE. Relationships of mortality with the fine and coarse fractions of long-term ambient PM_{10} concentrations in nonsmokers. *J Expo Anal Environ Epidemiol.* 2000;10(5):427-436. http://www.scientificintegrityinstitute.org/JEAEE090100.pdf

[d]Jerrett M, Burnett RT, Ma R, et al. Spatial Analysis of Air Pollution and Mortality in Los Angeles. *Epidemiology.* 2005;16(6):727-736. http://www.scientificintegrityinstitute.org/Jerrett110105.pdf

[e]Enstrom JE. Fine particulate air pollution and total mortality among elderly Californians, 1973-2002. *Inhal Toxicol.* 2005;17(14):803-816. http://www.arb.ca.gov/planning/gmerp/dec1plan/gmerp_comments/enstrom.pdf, and http://www.scientificintegrityinstitute.org/IT121505.pdf

[f]Enstrom JE. Response to "A Critique of 'Fine Particulate Air Pollution and Total Mortality Among Elderly Californians, 1973-2002" by Bert Brunekreef, PhD, and Gerard Hoek, PhD'. *Inhal Toxicol.* 2006:18:509-514. http://www.scientificintegrityinstitute.org/IT060106.pdf, and http://www.scientificintegrityinstitute.org/ITBH060106.pdf

[g]Zeger SL, Dominici F, McDermott A, Samet JM. Mortality in the Medicare Population and Chronic Exposure to Fine Particulate Air Pollution in Urban Centers (2000-2005). *Environ Health Perspect.* 2008;116:1614-1619. http://ehp03.niehs.nih.gov/article/info: doi/10.1289/ehp.11449

[h]Jerrett M. February 26, 2010 CARB Symposium Presentation by Principal Investigator, Michael Jerrett, UC Berkeley/CARB Proposal No. 2624-254 "Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort". 2010. http://www.scientificintegrityinstitute.org/CARBJerrett022610.pdf

[i]Jerrett M. October 28, 2011 Revised Final Report for Contract No. 06-332 to CARB Research Screening Committee, Principal Investigator Michael Jerrett, "Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort" Co-Investigators: Burnett RT, Pope CA III, Krewski D, Thurston G, Christakos G, Hughes E, Ross Z, Shi Y, Thun M. 2011. http://www.arb.ca.gov/research/rsc/10-28-11/item1dfr06-332.pdf, and http://www.scientificintegrityinstitute.org/Jerrett012510.pdf, and http://www.scientificintegrityinstitute.org/JerrettCriticism102811.pdf

[j]Lipsett MJ, Ostro BD, Reynolds P, et al. Long-term Exposure to Air Pollution and Cardiorespiratory Disease in the California Teachers Study Cohort. *Am J Respir Crit Care Med.* 2011;184(7):828-835. http://ajrccm.atsjournals.org/content/184/7/828.full.pdf

[k]Ostro B, Lipsett M, Reynolds P, et al. Long-Term Exposure to Constituents of Fine Particulate Air Pollution and Mortality: Results from the California Teachers Study. *Environ Health Perspect.* 2010;118(3):363-369. http://ehp03.niehs.nih.gov/article/info: doi/10.1289/ehp.0901181

[l]Jerrett M, Burnett RT, Beckerman BS, et al. Spatial analysis of air pollution and mortality in California. *Am J Respir Crit Care Med.* 2013;188(5):593-599. doi:10.1164/rccm.201303-0609OC. PMID:23805824.

[m]Ostro B, Hu J, Goldberg D, et al. Associations of Mortality with Long-Term Exposures to Fine and Ultrafine Particles, Species and Sources: Results from the California Teachers Study Cohort. *Environ Health Perspect.* 2015;123(6):549-556. http://ehp.niehs.nih.gov/1408565/, or http://dx.doi.org/10.1289/ehp.1408565

[n]Thurston GD, Ahn J, Cromar KR, et al. Ambient Particulate Matter Air Pollution Exposure and Mortality in the NIH-AARP Diet and Health Cohort. *Environ Health Perspect.* 2016;124(4):484-490. http://ehp.niehs.nih.gov/1509676/

US EPA. Regulatory Impact Analysis related to the Proposed Revisions to the National Ambient Air Quality Standards for Particulate Matter EPA-452/R-12-003. 2012. http://www.epa.gov/ttn/ecas/regdata/RIAs/PMRIACombinedFile_Bookmarked.pdf

Appendix C

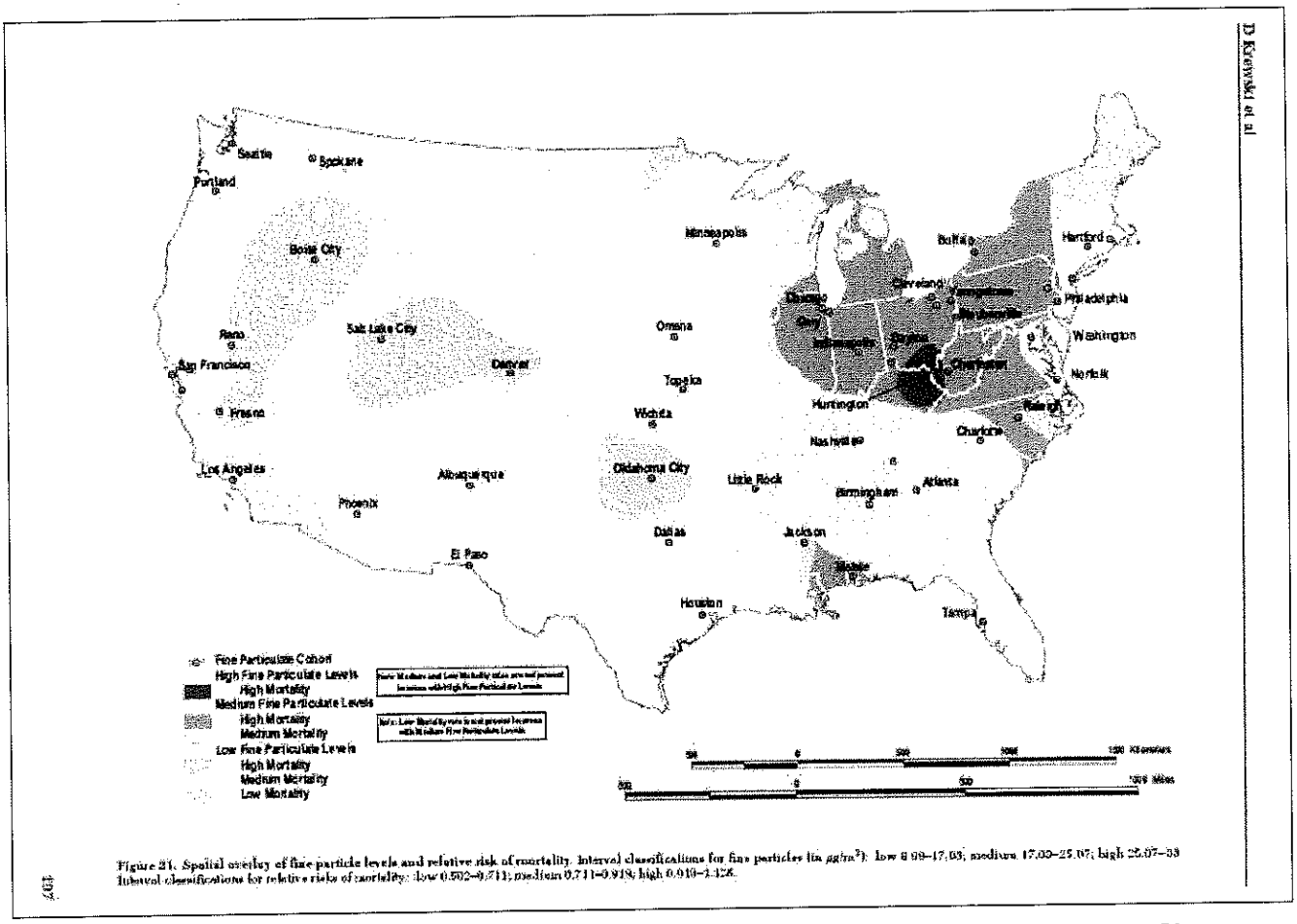


Figure C1. 1982 to 1989 $PM_{2.5}$ mortality risk (MR) in 50 cities (metropolitan areas) shown in Figure 21 on page 197 of HEI 2000[7,9] and listed in Appendix Table B1. Figure 21. Spatial overlay of fine particle levels and relative risk of mortality. Interval classifications for fine particles (in g/m^3): low 8.99 to 17.03: medium 17.03 to 25.07; high 25.07 to 33. Interval classifications for relative risks of mortality: low 0.052 to 0.711; medium 0.711 to 0.919; high 0.919 to 1.128.

Acknowledgments

The author thanks the American Cancer Society for helping initiate my epidemiologic career (http://www.scientificintegrityinstitute.org/Detels082773.pdf), for providing me with essential research support for many years (http://www.scientificintegrityinstitute.org/Mormon-LAT120689.pdf), for granting me unique access to California CPS I cohort data (http://www.scientificintegrityinstitute.org/CACP-SI090391.pdf), for selecting me as a Researcher who enrolled CPS II participants and worked with CPS II epidemiologists (http://www.scientificintegrityinstitute.org/Enstrom090213.pdf), and for making it possible for me to obtain unique access to the CPS II cohort data and detailed documentation. In addition, the author sincerely thanks Professors Melvin Schwartz, Lester Breslow, and Nikolai Vavilov, as well as Mr. Lehman Feldenstein, for the training and inspiration that made it possible for me to conduct and publish this research (http://www.scientificintegrityinstitute.org/AFAJEEAS051715.pdf).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The American Cancer Society provided the funding for the establishment of the CSP II cohort in 1982, the mortality follow-up from 1982 through 1988, and the preparation of the computerized files and documentation used for this research.

Supplemental Material

The online supplemental material is available at http://journals.sagepub.com/doi/suppl/10.1177/1559325817693345.

References

1. Pope CA III, Thun MJ, Namboodiri MM, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am J Resp Crit Care Med*. 1995;151(3 pt 1): 669-674. doi:10.1164/ajrccm.151.3.7881654.
2. Kaiser J. Showdown over clean air science. *Science*. 1997; 277(5325):466-469.

3. Gamble JF. $PM_{2.5}$ and mortality in long-term prospective cohort studies: cause-effect or statistical associations? *Environ Health Perspect*. 1998;106(9):535-549. doi:10.1289/ehp.98106535.
4. Phalen RF. The particulate air pollution controversy. *Nonlinearity Biol Toxicol Med*. 2004;2(4):259-292. doi:10.1080/15401420490900245. Accessed February 20, 2017.
5. Enstrom JE, Young SS, Dunn JD, et al. Particulate Matter Does Not Cause Premature Deaths. August 17, 2015. https://www.nas.org/images/documents/$PM_{2.5}$.pdf Within Wood P. Concerns about National Academy of Sciences and Scientific Dissent. National Association of Scholars. December 15, 2015. https://www.nas.org/articles/nas_letter. Accessed February 20, 2017.
6. Enstrom JE. *EPA's Clean Power Plan and $PM_{2.5}$-related Co-Benefits. Tenth International Conference on Climate Change. Panel 8. Heartland Institute*. Washington, DC: 2015. http://climateconferences.heartland.org/james-enstrom-iccc10-panel-8/, and http://www.scientificintegrityinstitute.org/JEEICCC061115.pdf. Accessed February 20, 2017.
7. Krewski D, Burnett RT, Goldberg MS, et al. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: Special Report. Cambridge, MA: Health Effects Institute; 2000. Part I. Replication and Validation and Part II. Sensitivity Analyses, particularly Figure 5 on page 161, Figure 13 on page 89, and Figure 21 on page 197 and Appendix D and Appendix F. https://www.healtheffects.org/publication/reanalysis-harvard-six-cities-study-and-american-cancer-society-study-particulate-air. Accessed February 20, 2017.
8. Enstrom JE. Fine particulate air pollution and total mortality among elderly Californians, 1973-2002. *Inhal Toxicol*. 2005; 17(14):803-816. PMID:16282158. http://scientificintegrityinstitute.org/IT121505.pdf
9. Enstrom JE. Response to "A Critique of 'Fine Particulate Air Pollution and Total Mortality Among Elderly Californians, 1973-2002'" by Bert Brunekreef, PhD, and Gerald Hoek, PhD. *Inhal Toxicol*. 2006;18(7):509-514. http://scientificintegrityinstitute.org/IT060106.pdf
10. Enstrom JE. Particulate Matter is Not Killing Californians. Proceedings of the American Statistical Association 2012 Joint Statistical Meeting, Section on Risk Analysis, San Diego, CA: 2012: pages 2324-2336. https://www.amstat.org/meetings/jsm/2012/proceedings.cfm, and http://www.scientificintegrityinstitute.org/ASAS092812.pdf
11. July 22, 2013 US House Science Committee Final Request to EPA for ACS CPS II Data. https://science.house.gov/news/press-releases/committee-threatens-subpoena-epa-secret-science, and https://science.house.gov/sites/republicans.science.house.gov/files/documents/07-22-2013%20Smith%20and%20Stewart%20to%20McCarthy.pdf. Accessed February 20, 2017.
12. August 1, 2013 US House Science Committee Subpoena to EPA Requesting ACS CPS II Data. https://science.house.gov/news/press-releases/smith-subpoenas-epa-s-secret-science, and https://science.house.gov/sites/republicans.science.house.gov/files/documents/Subpoena%20link.pdf. Accessed February 20, 2017.
13. Brawley OW. August 19, 2013 ACS Brawley Letter to EPA Refusing to Cooperate with August 1, 2013 US House Science Committee Subpoena of ACS CPS II Data. http://www.scientificintegrityinstitute.org/Brawley081913.pdf. Accessed February 20, 2017.
14. Gapstur SP. September 20, 2013 ACS Letter to Enstrom Denying CPS II Collaboration as Proposed in September 16, 2013 Enstrom Email. http://www.scientificintegrityinstitute.org/GapsturEns092013.pdf. Accessed February 20, 2017.
15. Greenbaum D. October 4, 2013 HEI Response to September 26, 2013 Enstrom Email Declining to Conduct Special Analyses of ACS CPS II re 2000 HEI Reanalysis Report.http://scientificintegrityinstitute.org/Greenbaum100413.pdf. Accessed February 20, 2017.
16. Hinton DO, Sune JM, Suggs JC, Barnard WF. Inhalable Particulate Network Report: Operation and Data Summary (Mass Concentrations Only). Volume I. April 1979-December 1982. EPA-600/4-84-088a. Research Triangle Park, NC: U.S. Environmental Protection Agency, November 1984, particularly pages 102-160 of 210 total pages. http://nepis.epa.gov/Exe/ZyPDF.cgi/20015OU3.PDF?Dockey=20015OU3.PDF. Accessed February 20, 2017.
17. Hinton DO, Sune JM, Suggs JC, Barnard WF. Inhalable Particulate Network Report: Data Summary (Mass Concentrations Only). Volume III. January 1983-December 1984. EPA-600/4-86/019. Research Triangle Park, NC: U.S. Environmental Protection Agency; April 1986: particularly pages 51-80 of 227 total pages. http://nepis.epa.gov/Exe/ZyPDF.cgi/9101R4L8.PDF?Dockey=9101R4L8.PDF
18. SAS, PHREG and REGRESSION Procedures, SAS/STAT 9.4 User's Guide. Cary, NC: SAS Institute Inc. http://support.sas.com/documentation/94/index.html. Accessed February 20, 2017.
19. Centers for Disease Control. National Center for Health Statistics. 1980 CDC WONDER On-line Database, compiled from Compressed Mortality File CMF 1968-1988. http://wonder.cdc.gov/cmf-icd9.html. Assessed April 15, 2016.
20. Krewski D, Jerrett M, Burnett RT, et al. Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality. HEI Research Report 140, Health Effects Institute, Boston, MA: 2009, particularly Table 34. https://www.healtheffects.org/publication/extended-follow-and-spatial-analysis-american-cancer-society-study-linking-particulate. Accessed February 20, 2017.
21. Krewski D. August 31, 2010 Letter to HEI re Special Analysis of California Subjects Within ACS CPS II Cohort Based on 2009 HEI Research Report 140 Methodology. http://www.arb.ca.gov/research/health/pm-mort/HEI_Correspondence.pdf. Accessed February 20, 2017.
22. Lipfert FW, Malone RG, Daum ML, Mendell NR, Yang CC. A Statistical Study of the Macroepidemiology of Air Pollution and Total Mortality. Brookhaven National Laboratory. Upton, NY. Report No. BNL 52122, April 1988, 136 pages. http://www.osti.gov/scitech/servlets/purl/7028097. Accessed February 20, 2017.
23. Lipfert F. Commentary on the HEI reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. *J Toxicol Environ Health A*. 2003;66(16-19):1705-1714; discussion 1715-1722. doi:10.1080/15287390306443.
24. Enstrom JE. July 11, 2008 CARB $PM_{2.5}$ Premature Deaths Teleconference Involving Enstrom, Pope, Jerrett, and

Burnett. Transcript and Audio File. http://www.scientificintegrityinstitute.org/CARB071108.pdf. Accessed February 20, 2017.

25. Enstrom JE. Analysis of HEI 2000 Figures 5 and 21 to Identify $PM_{2.5}$ Mortality Risk in 49 US Cities Used in Pope 1995 and HEI 2000. September 30, 2010. http://www.scientificintegrityinstitute.org/HEIFigure5093010.pdf
26. Enstrom JE. Submission to UCLA Research Integrity Officer Karagozian Challenging Jerrett et al. $PM_{2.5}$ Mortality Findings and Karagozian Response. December 19, 2016. http://scientificintegrityinstitute.org/RIOJerrettAll121916.pdf
27. Enstrom Compilation of Rejections of This Paper by Seven Major Journals. http://www.scientificintegrityinstitute.org/CPSIIRej122716.pdf. Accessed February 20, 2017.

EXHIBIT E

Risk Analysis, Vol. 36, No. 9, 2016

DOI: 10.1111/risa.12517

Inconsistencies in Risk Analyses for Ambient Air Pollutant Regulations

Anne E. Smith*

This article describes inconsistencies between health risk analyses that the U.S. Environmental Protection Agency (EPA) uses to support its decisions on primary National Ambient Air Quality Standards (NAAQS), and in the associated Regulatory Impact Analyses (RIAs) that accompany each NAAQS rulemaking. Quantitative risk estimates are prepared during the NAAQS-setting deliberations using inputs derived from statistical associations between measured pollutant concentrations and health effects. The resulting risk estimates are not directly used to set a NAAQS, but incorporated into a broader evidence-based rationale for the standard that is intended to demonstrate conformity with the statutory requirement that primary NAAQS protect the public health with a margin of safety. In a separate process, EPA staff rely on the same risk calculations to prepare estimates of the benefits of the rule that are reported in its RIA for the standard. Although NAAQS rules and their RIAs are released simultaneously, the rationales used to set the NAAQS have become inconsistent with their RIAs' estimates of benefits, with very large fractions of RIAs' risk-reduction estimates being attributed to populations living in areas that will already be attaining the respective NAAQS. This article explains the source of this inconsistency and provides a quantitative example based on the 2012 revision of the fine particulate matter ($PM_{2.5}$) primary NAAQS. This article also demonstrates how this inconsistency is amplified when criteria pollutant co-benefits are calculated in RIAs for non-NAAQS rules, using quantitative examples from the 2011 Mercury and Air Toxics Standards and the currently proposed Clean Power Plan.

KEY WORDS: Benefits; co-benefits; NAAQS; ozone; $PM_{2.5}$; regulatory impact analysis

1. BACKGROUND

When the primary particulate matter ($PM_{2.5}$) National Ambient Air Quality Standards (NAAQS) were first established in 1997 (one for annual average and one for daily average ambient $PM_{2.5}$ concentrations), the principal basis for those standards was epidemiological evidence of positive statistical associations between ambient $PM_{2.5}$ levels and adverse health effects, including premature death risk. These reported associations, combined with a presumption that they represented a causal relationship, were also used to calculate quantitative public health risk estimates to supplement reasoning on setting the NAAQS. Quantitative risk analyses based on epidemiological evidence have continued to be a central feature of the review process for revisions of the $PM_{2.5}$ NAAQS since then, and have also been a salient consideration in revisions of the NAAQS for ozone. This article focuses on a quantitative inconsistency that has emerged between the rationale that U.S. Environmental Protection Agency (EPA) Administrators use for setting a NAAQS when relying primarily on epidemiologically-based health risk evidence, and the estimates of public health benefits

*Address correspondence to Anne E. Smith, Senior Vice President, NERA Economic Consulting, 1255 23rd Street NW, Suite 600, Washington, DC 20037, USA; anne.smith@nera.com.

from those rules that EPA staff produces in its Regulatory Impact Analyses (RIAs).[1]

2. THE RATIONALE FOR SETTING A PRIMARY NAAQS

The Clean Air Act requires EPA[2] to set the primary NAAQS for each criteria pollutant at levels that "are requisite to protect the public health" while "allowing an adequate margin of safety."[1] This determination must be made without regard to the potential cost of meeting the standard,[2] and legal rationales for choosing a NAAQS traditionally involved a balanced consideration of three attributes: (1) size of affected population, (2) severity of effect, and (3) certainty of effect.[3] However, the evolution since 1997 towards greater reliance on epidemiological evidence in setting a NAAQS forced a shift in how the rationale could be constructed, particularly for $PM_{2.5}$. This was because the available epidemiological studies on several clearly adverse types of health effects (such as premature death) have not been able to identify a "threshold" or any other less sharp delineation of a point where the risk per unit increment of concentration appears to attenuate.[3] This situation eliminates the first two of the three above-mentioned considerations that EPA had typically relied on in NAAQS-setting rationales. That is, (1) the entire U.S. population is now implicated as at risk at every potential NAAQS level, and (2) the severity of effect can no longer be seen to be changing as lower potential NAAQS levels are considered. As a result, consideration (3)—uncertainty about the reliability of the epidemiologically estimated association—has become the only consideration remaining available to EPA for setting a primary NAAQS above zero that can be argued to be adequately protective of the public health as required by the statute.

This shift in the nature of the scientific evidence for setting a NAAQS was so profound that the U.S. Court of Appeals ruled that the setting of a NAAQS under these circumstances amounted to an unconstitutional delegation of legislative power to the Administrator unless she would first articulate an "intelligible principle" for how to draw that line.[6] However, the Supreme Court overruled this finding,[7] with the result being that since then EPA's rationales for at least two of the NAAQS (i.e., $PM_{2.5}$ and ozone) have largely emphasized identifying a level at which continuation of the nonthreshold statistical health associations becomes too uncertain to indicate an actionable level of further public health risk.

The preamble for the 2012 $PM_{2.5}$ NAAQS decision provides an example. It starts by noting that setting a standard based on epidemiological studies that cannot identify a population threshold requires a decision-making approach that "includes consideration of how to weigh the uncertainties in the reported associations across the distributions of $PM_{2.5}$ concentrations in the studies and the uncertainties in quantitative estimates of risk, in the context of the entire body of evidence before the Agency."[8] Later, the document states, "[i]n reaching decisions on alternative standard levels to propose, the Administrator judged that it was most appropriate to examine where the evidence of associations observed in the epidemiological studies was strongest and, conversely, where she had appreciably less confidence in the associations observed in the epidemiological studies,"[9] and after a detailed discussion of the epidemiological information states, "[t]he Administrator views this information as helpful in guiding her determination as to *where her confidence in the magnitude and significance of the associations is reduced to such a degree* [emphasis added] that a standard set at a lower level would not be warranted to provide requisite protection that is neither more nor

[1] A separate point of discussion regarding the quantitative risk estimates is whether the full body of scientific evidence is sufficient to give confidence that these epidemiological associations reflect a causal relationship between the pollutant and health endpoint studied. This article does not attempt to add to that discussion.

[2] Formally, under the Clean Air Act, the responsibility for deciding where to set a NAAQS is vested specifically in the Administrator. Throughout this article, when I use the term "EPA," I am referring to the EPA Administrator. When not referring to the Administrator specifically, I use the terms "EPA staff" or "Agency."

[3] EPA staff and others often refer to this as a "threshold" for effects, but the phenomenon being sought to help identify a protective level for a particular adverse effect need not be a point of sharp delineation where all population-wide effects end. Even evidence of diminishment in the slope of the association would be helpful but has not been consistently found. Lack of detection of such a diminishment in an association, even if the detected association is causal at relatively high concentrations, does not mean one does not exist at some relatively low concentration (see Ref. 4, p. 382). This is because the epidemiological techniques available have very limited ability to reliably discern the shape of a potential concentration-response relationship, and thus to inform the question of where or whether the association may end. It is theoretically established that unavoidable inaccuracies in measurement of an explanatory variable (e.g., pollutant exposure) make it difficult to statistically detect a threshold or other nonlinearity at low concentrations even when it actually exists.[5]

less than needed to provide an adequate margin of safety."[10]

Similarly, in 2008 EPA used lack of confidence in continuation of the epidemiological associations to lower levels as its rationale for not setting the ozone NAAQS lower than 0.075 ppm despite clinical evidence in the record of health responses at yet lower concentrations. The ozone NAAQS preamble states: "A standard set at a level lower than 0.075 would only result in significant further public health protection if, in fact, there is a continuum of health risks in areas with 8-hour average O_3 concentrations that are well below the concentrations observed in the key controlled human exposure studies and if the reported associations observed in epidemiological studies are, in fact, causally related to O_3 at those lower levels. Based on the available evidence, the Administrator is not prepared to make these assumptions. Taking into account the uncertainties that remain in interpreting the evidence from available controlled human exposure and epidemiological studies at very low levels, the Administrator notes that *the likelihood of obtaining benefits to public health with a standard set below 0.075 ppm O_3 decreases* [emphasis added], while the likelihood of requiring reductions in ambient concentrations that go beyond those that are needed to protect public health increases."[11] The U.S. Court of Appeals for the District of Columbia Circuit accepted this rationale and upheld the standard in 2013.[12]

Although the NAAQS rationales are not written to conform to the terminology of probability or expected values, readers with decision analytic or other risk analysis training would be inclined to interpret the above quotes as expressing subjective judgments about the probability that the health relationships apparent in statistical associations cease to exist at some point on the continuum of lower and lower ambient pollutant concentrations. A decision-analytic interpretation of the above statements might be as follows. In order for a selected NAAQS level to be deemed as requisite to protect the public health, EPA's subjective probability that the relationship exists at and below the selected NAAQS level must, logically, be very nearly zero. (Indeed, the subjective probability of continued effects must fall to nearly zero at an ambient concentration somewhere *above* the selected NAAQS level. This is because the NAAQS needs to include at least some margin of safety, and thus must be set at least somewhat lower than the level where expected risk is deemed to become too small to be considered a public health concern.)

3. THE RESULTING INCONSISTENCY IN BENEFITS ESTIMATES FOR A NAAQS

Thus, in setting NAAQS using epidemiological evidence, EPA has deemed quantitative estimates of health risks for concentrations below the NAAQS far less reliable and more inaccurate than the numerical precision with which they are reported. In essence, the NAAQS rationales give little or no weight to the subset of the quantitative risk estimates the Agency has placed in the record that have been calculated for pollutant concentrations below the selected NAAQS level. This lack of confidence in risk estimates from that below-the-NAAQS range does not, however, make its way into the RIAs that accompany the release of the final rules.

RIAs are documents that report on the benefits and costs of each major new regulation, such as a revised NAAQS. Federal regulatory agencies are required to prepare RIAs by Executive Order of the President.[13,14] Although this requirement is unrelated to the legal requirements of the statute that motivates the regulation (such as the Clean Air Act in the case of air pollutant regulations), EPA's RIAs for air regulations adopt the same epidemiologically-based method of quantifying health risks used when deliberating where to set the NAAQS.[4] The consistency ends there, however. At the same time that EPA is setting NAAQS at levels where it has minimal confidence that the public health is affected at lower concentrations, the Agency's RIAs are giving the same weight to risks calculated for population exposures *below* the NAAQS level as they do to risks calculated for population exposures above the NAAQS level. That is, RIAs assume elevated hazards exist with 100% certainty for all ambient pollutant exposure levels down to a zero concentration, inconsistent with EPA's judgments (formed when assessing those pollutants' hazards), which imply nearly 0% certainty. EPA does not explain or try to justify why data that are too uncertain to use in the NAAQS preamble context are certain enough to use in the RIA context. Although different certainty standards may be

[4]While the "benefits" in an RIA are stated as a monetary value to be compared to the regulation's costs, they are directly derived from quantitative estimates of physical health effects.

justified in the context of decisions with different consequences, the contexts of a NAAQS preamble and that NAAQS's RIA are not very different at all.

This inconsistency was not always as pronounced as it is now. Until 2009, risk reduction calculations used in air RIAs were at least truncated for pollutant concentrations below the lowest concentration level measured in the epidemiological study being used to make the risk estimates. RIAs would still include risk reduction estimates below the prevailing NAAQS level, as NAAQS levels have always been set at levels above the lowest levels measured in the studies. However, from 2009 onwards, RIAs eliminated even that truncation, which resulted in a sudden and large increase in RIA benefits estimates for $PM_{2.5}$ and ozone pollutant changes.[15] The fact that RIAs calculate health risk reductions below the NAAQS, and now down to zero, is widely known but the following examples quantify the extent to which this practice results in upward-biased risk and benefits estimates. This author recommends that EPA staff more clearly communicate subjective epistemic uncertainty in its RIA benefits estimates. More specifically, the author recommends that the Agency's central estimates of benefits in its RIA be made consistent with the science-policy judgments EPA makes in setting the criteria pollutant standards. This recommendation is in line with the need for more effective sensitivity analysis capabilities for health risk analyses, as described by Smith and Gans.[16]

4. OVERSTATEMENT OF EXPECTED BENEFITS OF THE 2012 $PM_{2.5}$ PRIMARY NAAQS REVISION

The implications of this inconsistency are illustrated using as an example the RIA for the 2012 PM NAAQS rulemaking.[17] In this rulemaking, the annual primary standard for $PM_{2.5}$ was tightened from an annual average of 15 to 12 $\mu g/m^3$. In the associated RIA, a range of 460 to 1,000 fewer premature deaths per year was estimated from tightening the standard to 12 $\mu g/m^3$. This range was derived by applying two different concentration-response functions to the Agency's standard risk calculation formula. The concentration-response coefficient for the lower end of the range was derived using a coefficient from Krewski *et al.*,[18] and the upper end of the range was derived using a coefficient from Lepeule *et al.*[19] A yet wider range of uncertainty in potential mortality risk reductions exists, as explained in Ref. 16, but the following discussion addresses only how the Agency's own range changes when the assumptions of the RIA's risk analysis are made consistent with EPA's reasoning when choosing how stringently to set the standard.

Calculations were performed using EPA's BenMAP model, which is a PC-based program that enables users to compute health risks associated with criteria pollutants using the standard formulas that EPA uses in its own RIAs, and using EPA's or their own input files and other assumptions.[20] The air quality input files that had been used for this RIA's calculations were obtained from EPA staff. After confirming that BenMAP does indeed replicate the mortality reduction estimates reported in the RIA using those data, the same files were then used to assess the portion of the RIA's premature mortality estimates that are associated with the linear, no-threshold assumption that assumes that the risk relationship continues to exist below the selected NAAQS. This analysis found that 70% of the benefits for the standard of 12 $\mu g/m^3$ were due to reductions in $PM_{2.5}$ from baseline levels that were already attaining (i.e., lower than) that standard.

Given that the choice of a NAAQS level of 12 $\mu g/m^3$ meant that EPA assigned too little confidence in the continuation of health effects below 12 $\mu g/m^3$ to warrant setting the NAAQS at a lower level, standard decision analysis would assign negligible probability to calculations of benefits from reductions that would be occurring from levels below that NAAQS. That is, the *expected* values for 70% of the Agency's risk calculations should be approximately zero. When a threshold is assumed at 12 $\mu g/m^3$, BenMAP calculates that the expected risk reduction of that NAAQS would be 138 to 313 fewer premature deaths per year, considerably lower than the 460 to 1,000 deaths reported in the RIA. (Dollar values of the benefits also fall proportionally.)

As noted above, the rationale for the NAAQS arguably implies that some of the benefits derived from locations with concentrations just above 12 $\mu g/m^3$ also should be given less than 100% weight because of EPA's assurance that exposures to annual average concentrations of 12 $\mu g/m^3$ are protective *with an adequate margin of safety*. EPA rarely if ever defines the magnitude of its margin of safety quantitatively. However, ranges for its magnitude could be tested with sensitivity analyses. If, for example, the margin of safety is taken to be about 1 $\mu g/m^3$, and a threshold is assumed in the risk relationship 13 $\mu g/m^3$, BenMAP calculates the expected benefits associated with the selected NAAQS of 12 $\mu g/m^3$ are

Table I. Estimates of Avoided Premature Deaths in 2020 for the 12 $\mu g/m^3$ $PM_{2.5}$ NAAQS: RIA Assumptions Compared to Alternative Views Suggested by EPA's Rationale for that NAAQS

Confidence Category (baseline $PM_{2.5}$ concentration)	NAAQS-Based Risk Reduction Estimate	RIA-Based Risk Reduction Estimate (% of total)
Already attaining ($\leq$12 $\mu g/m^3$)	Approximately 0	318 (70%)
Not attaining/in margin (e.g., >12 to 13 $\mu g/m^3$)	0–117	117 (26%)
Not attaining/above margin (e.g., >13 $\mu g/m^3$)	21	21 (5%)
Confidence weighted		
Total risk reduction estimate	21–117	456

only 21 to 48 deaths, less than 5% of the RIA's estimate of benefits from that standard.

Whether the particular assumptions in this analysis about where the concentration-response relationship begins to exist are reasonable or should be refined, its point is that the RIA's benefits estimates are very sensitive in the downward direction to expressions of declining confidence in continuation of the association at or just above the selected NAAQS level. The result is that the RIA benefits are substantially overstated compared to those that would more appropriately reflect the subjective weights expressed by EPA in its rationale for setting the standard at 12 $\mu g/m^3$. Table I contrasts the results of the RIA with judgments about confidence in those risk calculations that one might infer from the NAAQS rationale, and illustrates one way that RIAs could be enhanced to better communicate to the public the implications of the judgments made in setting the NAAQS for the rule's benefits estimates.

For simplicity, Table I summarizes only the lower-bound benefits estimate of 460 deaths (which BenMAP calculates more precisely as 456 deaths).[5] In this table, the risk estimates are divided into three "confidence categories." The lowest confidence category is for risk reductions attributed to populations already residing in areas of attainment (i.e., with annual average concentrations less than 12 $\mu g/m^3$). Given the NAAQS rationale, the public health risk is *de minimis*, and in weighted terms, would be nearly zero, while in the RIA, which gives 100% weight to all such risk calculations, benefits equal to about 318 deaths per year are assigned. The middle confidence category is for risk reductions attributed to populations in areas that are just above the NAAQS before the standard is implemented, but close enough to attainment that they might be viewed as being within the (undefined) "margin of safety." (For purposes of constructing the illustrative tabular summary, the margin of safety is assumed to be about 1 $\mu g/m^3$, meaning that less than the NAAQS-based weights would be declining or perhaps nearly zero even within this category of baseline exposures.) To reflect risk estimates that fall in this category, the NAAQS-based risk reduction estimate is listed as being somewhere between 0 and 117, while the RIA would assign it 117 with 100% confidence.

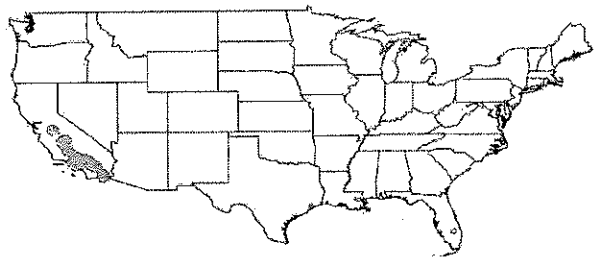


Fig. 1. Areas projected in the $PM_{2.5}$ NAAQS RIA to experience health benefits under the selected NAAQS of 12 $\mu g/m^3$ (456–1,033 avoided premature deaths, rounded to nearest death).

Finally, there are 21 avoided premature deaths estimated for populations living in areas well above the NAAQS. For this third category, the RIA's benefits estimates can be considered consistent with the NAAQS-based rationale. Note that for the $PM_{2.5}$ NAAQS RIA, this category accounts for only about 5% of the total RIA benefits estimate. It is recommended that RIAs provide their benefits estimates for criteria pollutants in a format such as Table I, and more explicitly provide weighted benefits estimates for confidence categories that are defined *with respect to the NAAQS level.*

Geographical representation of where these health benefits are expected to occur is also interesting to explore. The $PM_{2.5}$ NAAQS RIA calculated reductions in premature mortality only for areas that

[5]The upper-bound risk estimates would fall into the three rows in the table in the same proportions as seen for the lower-bound estimates in the table.

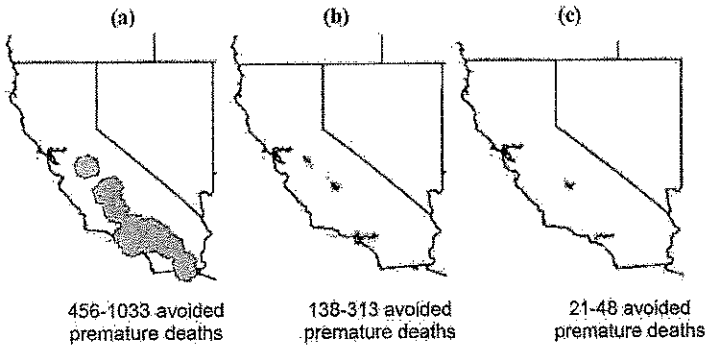


Fig. 2. Sensitivity analysis of areas projected to experience health benefits under the 12 μg/m3 NAAQS: (a) assuming benefits for all baseline PM2.5 levels; (b) assuming risks exist only if baseline PM2.5 is above 12 μg/m3; (c) assuming risks exist only if baseline PM2.5 exceeds the selected standard by more than 1 μg/m3.

were within 50 km of a monitor that the RIA's air quality analysis projected would not attain the new standard under baseline conditions. Fig. 1 shows the locations in which the RIA's estimate of 460–1,000 avoided premature deaths occur. It is notable that all of those benefits occur in California. Fig. 2 zooms in on California to show: (a) the areas in Fig. 1 where benefits are attributed to reductions in $PM_{2.5}$ at any level (i.e., showing the same areas as in Fig. 1); (b) the more limited areas projected to experience a health benefit when only reductions in $PM_{2.5}$ that start above the 12 $\mu g/m^3$ NAAQS are considered; and (c) the even more limited areas if a 1 $\mu g/m^3$ margin of safety is assumed to be associated with the selected standard of 12 $\mu g/m^3$. That is, Fig. 2(c) only gives weight to risks below 13 $\mu g/m^3$. Both Figs. 2(b) and (c) reveal a far smaller area of at-risk populations than assumed in the RIA (i.e., than in Fig. 2(a)).

This example from the $PM_{2.5}$ NAAQS RIA brings to light another important uncertainty in its mortality benefits. All of the benefits estimates for the NAAQS of 12 $\mu g/m^3$ are based on $PM_{2.5}$ changes in California. The risk calculations for changes in $PM_{2.5}$ in California are performed using relative risk estimates derived from the entire United States, yet the epidemiological evidence that an association between $PM_{2.5}$ and all-cause mortality risk exists in California is tenuous.[6] Hence all of the above risk estimates might actually be zero, even if one does not wish to discount risks in areas already below the NAAQS. In other words, the much tighter 2012 $PM_{2.5}$ NAAQS was set on the basis of projected mortality reductions that occur only in a part of the United States where the evidence of heightened mortality risk from $PM_{2.5}$ appears to be weaker than in other parts of the United States.

5. OVERSTATEMENT OF CRITERIA POLLUTANT CO-BENEFITS IN NON-NAAQS RULEMAKINGS

As explained in Ref. 15, epidemiologically-based estimates of co-benefits from coincidental reductions of ambient criteria pollutants (especially $PM_{2.5}$) have also driven statements about regulatory benefits for a majority of non-NAAQS air rulemakings in recent years. The upward bias in RIA benefits estimates becomes even more pronounced when co-benefits are calculated from coincidental criteria pollutant reductions under regulations that do not relate to the NAAQS or regulations to help attain a NAAQS. Prominent examples are the RIAs for the Mercury and Air Toxics Standards (MATS) for electricity-generating units promulgated in December 2011[21] and the Clean Power Plan (CPP) proposed in June 2014.[22]

The MATS RIA projected $PM_{2.5}$ co-benefits in the hundreds of billions of dollars per year, based almost entirely on estimates of reduced premature mortality from reductions in $PM_{2.5}$: 4,200 to 11,000 deaths per year. The reductions in $PM_{2.5}$ in the MATS RIA are projected to occur when generating units are forced to install controls to reduce acid gas emissions, which will also reduce SO_2 emissions, a precursor to ambient $PM_{2.5}$ formation. A figure in the MATS RIA reveals that over 99% of those projected benefits are projected to occur in areas where the $PM_{2.5}$ levels will already be below the $PM_{2.5}$ NAAQS

[6]The $PM_{2.5}$ RIA[17] cites seven California-specific $PM_{2.5}$ cohort studies with all-cause risk estimates and notes that four have insignificant associations while three have larger coefficients (Ref. 17 at p. 5, A-13). However, one of the three positive findings cited (i.e., Ostro *et al.*, 2010) was erroneous, according to an erratum published the following year (Ostro *et al.*, 2011), and the corrected estimate of association was found to be insignificant. The remaining two positive findings cited were from the same cohort, one estimate being just an update of the other. Thus, the evidence for an all-cause mortality association in California alone consists of five null findings and one cohort with a positive finding.

of 12 $\mu g/m^3$ (Figure 5–15 on p. 5–102 of Ref. 21). If the MATS rule's co-benefits are calculated probabilistically, accounting for the very low subjective probability that EPA assigned to the existence of the $PM_{2.5}$-health effects relationships at levels below the NAAQS, the resulting estimate of expected benefits from the MATS rule becomes nearly zero.

The fraction of the $PM_{2.5}$ co-benefits calculated below the NAAQS is much higher in the MATS RIA than the already high level of 70% that we have found for the benefits calculated for the $PM_{2.5}$ NAAQS rule itself. This is due to the fact that benefits in the RIA for the NAAQS rule were calculated only in areas within 50 km of a monitor that was projected to be out of attainment. By letting projected nonattainment constrain the geographical area over which benefits will be calculated, one ensures that a larger fraction of the resulting benefits will indeed be from areas above the NAAQS. However, when co-benefits of some other rule are assessed using $PM_{2.5}$ risk relationships, no such constraint is applied. In the MATS rule, co-benefits were calculated across the entire nation, and furthermore, the units where acid gas controls were incremental to baseline controls were more likely to be in areas already attaining the NAAQS. As a result, nearly all of the $PM_{2.5}$ co-benefits are projected in NAAQS-attaining areas. For these reasons, the bias in $PM_{2.5}$ co-benefits estimates in RIAs for non-$PM_{2.5}$ rulemakings will tend to be much greater than the bias in the direct benefits estimates in RIAs for $PM_{2.5}$ regulations.

The same magnitude of overstatement of co-benefits is apparent in the RIA for the proposed CPP RIA, which includes co-benefits for both $PM_{2.5}$ and ozone. In the CPP RIA (focusing, for simplicity, on its Option 1 with state-level implementation) the $PM_{2.5}$ co-benefits of the rule are estimated to be up to 4,100 deaths in 2020 and up to 6,200 deaths in 2030, and the ozone co-benefits are estimated to be up to 170 and 440 in those respective years (Tables 4–16 through 4–18 on pp. 4–34 to 4–36 of Ref. 22). Unlike the MATS RIA, the CPP RIA does not provide any information on the fraction of these co-benefits that are calculated for areas already attaining those two NAAQS, but they can be inferred by replicating the co-benefits calculations from other data in the RIA.[7] Recalling that the $PM_{2.5}$ NAAQS RIA indicates that only California will be exceeding the $PM_{2.5}$ NAAQS in 2020, only California-based $PM_{2.5}$ co-benefits estimates could be associated with exposures in the above-the-NAAQS category: less than 1% of the CPP RIA's $PM_{2.5}$ co-benefits are attributable to changes in emissions in California in 2020. Furthermore, the $PM_{2.5}$ NAAQS is supposed to be fully attained by 2020, so even that sliver of the $PM_{2.5}$ co-benefits attributable to California are supposedly in an attainment area. Although California is not projected to attain the ozone NAAQS before 2030, less than 0.5% of the ozone-related co-benefits are associated with changes in ozone precursors in California. Thus, in the CPP RIA as well in the MATS RIA, more than 99% of the co-benefits would be discounted if health risks below the NAAQS are assigned a much lower probability (or confidence weight) than risks above the NAAQS.

[7]This involves using data on emissions reductions of the $PM_{2.5}$ and ozone precursor emissions in the RIA's Table 4–10, and multiplying them by the incidence-per-ton estimates in Tables 4A-5 through 4A-7.

6. CONCLUSION

In conclusion, we find that a large majority of the Agency's estimated health benefit from the 2012 $PM_{2.5}$ NAAQS are attributable to reductions of $PM_{2.5}$ in areas that are already in attainment of the $PM_{2.5}$ NAAQS. RIA calculations of risk reduction in areas already attaining the new NAAQS are given the same weight (i.e., subjective confidence level) as projected benefits from areas that would be exceeding the NAAQS. These RIA calculations are based on assumptions that are inconsistent with the rationale for that NAAQS. The above sensitivity analyses show that this causes RIAs' benefits estimates to be much larger than estimates of the expected benefits that can be reasonably inferred from EPA's NAAQS-setting rationale. The overstatement becomes nearly 100% for co-benefits from criteria pollutants in RIAs for non-NAAQS regulations, such as the MATS rule and the proposed CPP rule. RIAs should be written to reflect consistency with EPA's NAAQS policy judgments. Precise confidence weights will likely never be articulated, but this article has shown that the quantitative importance of such policy judgments for benefits estimates can be communicated to RIA readers in simple formats. It is the opinion of this author that such quantitative disclosure is important to maintaining credibility and trust in the Agency's RIAs.

ACKNOWLEDGMENTS

This work was conducted with funding from the Electric Power Research Institute. The author thanks Ms. Reshma Patel for her analytical support

in preparing the analyses used in this article. The author also thanks three anonymous reviewers for their comments and suggestions. Any errors remain the author's sole responsibility.

REFERENCES

1. Section 109 (b)(1), Clean Air Act; 42 USC §7409.
2. Whitman v. American Trucking Associations, 531 U.S. 473–476 (2001).
3. Lead Industries Association Inc v. Environmental Protection Agency, 647 F.2d 1130 (1980).
4. Smith AE. Response to commentary by Fann *et al.* on "Enhancing the characterization of epistemic uncertainties in $PM_{2.5}$ risk analyses." Risk Analysis, 2015; 35(3):381–384.
5. Brauer M. Brumm J. Vedal S. Petkau AJ. Exposure misclassification and threshold concentrations in time series analyses of air pollution health effects. Risk Analysis, 2002; 22(5):1183–1193.
6. American Trucking Associations v. EPA, 175 F. 3d 1027, 1034–1037 (D.C. Cir. 1999), rehearing granted in part and denied in part, 195 F. 3d 4 (D.C. Cir. 1999), affirmed in part and reversed in part. Whitman v. American Trucking Associations, 531 U.S. 457 (2001).
7. Whitman v. American Trucking Associations, 531 U.S. 473–476 (2001).
8. 78 Fed. Reg. 3086 at 3098.
9. 78 Fed. Reg. 3086 at 3139.
10. 78 Fed. Reg. 3086 at 3161.
11. 76 Fed. Reg. 16436 at 16483.
12. Mississippi v. EPA, 744 F.3d 1334 (D.C. Cir. 2013).
13. Executive Order 12866. 1993. "Regulatory Planning and Review." 58 Fed. Reg. 51735. October 4. Available at: www.whitehouse.gov/omb/inforeg/eo12866.pdf.
14. Executive Order 13563. 2011. "Improving Regulation and Regulatory Review." 76 Fed. Reg. 3821. January 18. Available at: http://www.regulations.gov/exchange/sites/default/files/doc_files/President%27s%20Executive%20Order%2013563_0.pdf.
15. Smith, AE. An evaluation of the $PM_{2.5}$ health benefits estimates in regulatory impact analyses for recent air regulations. Report prepared for the Utility Air Regulatory Group, December 2011, at p. 24. Available at: http://www.nera.com/content/dam/nera/publications/archive2/PUB_Smith_Ozone NAAQS_0711.pdf.
16. Smith AE, Gans W. Enhancing the characterization of epistemic uncertainties in $PM_{2.5}$ risk analyses. Risk Analysis, 2015: 35(3):361–78.
17. EPA. Regulatory impact analysis for the final revisions to the National Ambient Air Quality Standards for particulate matter, EPA-452/R-12-003, December, 2012.
18. Krewski D. Jerrett M. Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA 3rd, Thurston G, Calle EE, Thun MJ, Beckerman B. DeLuca P, Finkelstein N, Ito K. Moore DK, Newbold KB, Ramsay T. Ross Z. Shin H. Tempalski B. Extended follow-up and spatial analysis of the American Cancer Society Study linking particulate air pollution and mortality. Research Report Health Effects Institute, 2009; 140:5–114.
19. Lepeule J, Laden F, Dockery D, Schwartz J. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard Six Cities Study from 1974 to 2009. Environmental Health Perspectives, 2012; 32(1):81–95.
20. BenMAP version 4.0.67 Available at http://www.epa.gov/air/benmap/download.html.
21. EPA. Regulatory impact analysis for the final mercury and air toxics standards. EPA-452/R-11-011, December 2011.
22. EPA. Regulatory impact analysis for the proposed carbon pollution guidelines for existing power plants and emission standards for modified and reconstructed power plants. EPA -542/R-14-002, June 2014.

EXHIBIT F

Risk Analysis, Vol. 36, No. 9, 2016

DOI: 10.1111/risa.12670

Invited Commentary

Rethinking the Meaning of Concentration–Response Functions and the Estimated Burden of Adverse Health Effects Attributed to Exposure Concentrations

Louis Anthony (Tony) Cox Jr.*

Four articles by Anenberg *et al.*, Fann *et al.*, Shin *et al.*, and Smith contribute valuable perspectives and syntheses to a large and growing literature that estimates the burden of mortality risks attributed to fine particulate matter (PM2.5) based on estimated epidemiological associations, summarized as concentration–response (C–R) relations. This comment questions the use of C–R relations to predict or estimate how changing exposure concentrations would change responses in a population. C–R associations typically reflect modeling choices, and equally good choices can commonly lead to conflicting conclusions about the signs, significance, and magnitudes of C–R relations and regression coefficients. This indicates that C–R relations do not necessarily reflect underlying stable causal laws useful for making risk predictions, but only choices about how to describe past data, with no uniquely correct choice being determined by the data. Similarly, currently available C–R data typically do not suffice to make valid predictions about how future changes in concentrations will affect responses. These difficulties can be substantially overcome by model ensemble and causal graph modeling and time series methods, but these require different data and knowledge—for example, knowledge of how multiple variables depend on each other, rather than only of how one dependent variable is associated with multiple explanatory variables—than that captured by traditional C–R models or expressible by any single C–R coefficient or curve.

KEY WORDS: Air pollution; ambiguous association; causality; causation; concentration–response function; fine particulate matter

1. INTRODUCTION

To what extent can historical exposure–response (C–R) associations be used to predict correctly how future changes in exposure would affect responses? A voluminous literature, including considerable authoritative expert opinion, assumes that observed C–R relations and associations can be used to give useful guidance to policymakers about how changes in concentrations should be expected to change health responses in populations. If the C–R model is some regression curve or function *response* = *f*(*concentration*), then this view holds that changing the concentration from old level x to new level y should change the population response from $f(x)$ to $f(y)$. The purpose of this comment is to dispute that view. An excellent—indeed, even a perfect—descriptive model of the relation between past levels of exposure concentrations and responses does not necessarily or usually allow us to predict how changing concentration would change responses.

C–R functions estimated from past data are widely used to estimate the human health burdens of different exposures and to project how changes in exposures would change health impacts. For example, in this issue of *Risk Analysis*, Shin *et al.*[1] mention that "[o]ur approach to characterizing the shape of the exposure–response function is based on only summary information available in the open literature: relative risk estimates and the exposure

*Cox Associates and University of Colorado; tcoxdenver@aol.com.

distribution for each study." They suggest that their approach based on statistical associations (relative risk estimates) is "more useful for burden estimation" and can be used to "present an example of mortality risk due to long-term exposure to ambient fine particulate matter," reflecting a belief that relative risks provide a useful basis for estimating disease burdens and mortality risks due to (presumably meaning caused by, and reducible by reducing) exposure concentrations. Similarly, Fann *et al.*[2] state: "At the core of these assessments are judgments about the likelihood that PM2.5 is a causal factor in mortality and about the choices made to characterize the C–R function that quantified the relationship between changes in concentrations of ambient PM2.5 and the risk of premature mortality. ... As a final evaluation, we examine the implications of any differences among the quantitative methods for estimating the number of avoided PM2.5-related premature deaths, including uncertainty, for an illustrative policy scenario." Again, this appears to reflect a belief that C–R functions can tell us how "changes in concentrations of ambient PM2.5" would affect "risk of premature mortality" and "the number of avoided PM2.5-related premature deaths." Smith notes that "[q]uantitative risk estimates are prepared during the NAAQS-setting deliberations using inputs derived from statistical associations between measured pollutant concentrations and health effects. The resulting risk estimates are not directly used to set a NAAQS, but incorporated into a broader evidence-based rationale for the standard that is intended to demonstrate conformity with the statutory requirement that primary NAAQS protect the public health with a margin of safety."[3] Anenberg *et al.*[4] state that their article "reviews 12 current and publicly available multinational tools that combine air quality information, epidemiologically–derived concentration–response associations, and demographic data sets to estimate air-pollution-related health risks." These descriptions correctly reflect the reality that current quantitative risk estimates used to inform regulatory policy deliberations about health harms caused by exposures to pollutants are based on epidemiologically-derived statistical associations (e.g., C–R regression coefficients and relative risks or odds ratios), perhaps augmented with expert judgments about whether exposure is "a causal factor," rather than on quantification and validation of direct and indirect causal impacts of changes in concentrations on changes in responses.

The purpose of this comment is to challenge the belief that statistical associations and C–R relations between historically observed exposure concentrations and responses provide the information needed to draw valid causal inferences about how changing exposure concentrations would change responses. This typically cannot be determined from data on past C–R associations, even with the help of expert judgments. Instead, such inferences require understanding how changes in causes will change effects. This requires data, knowledge, and analyses different from those in C–R functions. To fix ideas and to concretely illustrate methodological points, we use both simple examples and a publicly available data set from the Los Angeles air basin, kindly provided by Dr. Stanley Young. The data, and statistical tools for analyzing it, are included in the free Causal Analytics Toolkit (CAT) software described at https://regulatorystudies.columbian.gwu.edu/causal-analytics-toolkit-cat.

1.1. Which C–R Relation? Signs, Magnitudes, and Significance of Associations Depend on Modeling Choices

When an associational analysis of a data set, such as a regression model or a relative risk, odds ratio, or slope factor calculation, reveals a significant positive C–R relation, it is tempting to think that one has thereby learned something about the real world: that higher concentrations are associated with higher response rates. But this natural interpretation is often mistaken. Usually, a significant positive association shows only that the investigator has selected a model that gives a significant positive C–R association (e.g., slope or regression coefficient) when applied to the data set. Selecting different, equally good (or better) models (by any criterion) for the same data might produce no significant positive association, or a significant negative association. This poses a methodological challenge, recognized by Dominici *et al.*,[5] who noted that associational methods are unreliable in general, as their results can be reversed by making different modeling choices.

Table I illustrates an extreme case of this point using a simple hypothetical example for three communities, A, B, and C, having different concentrations of PM2.5 and different elderly mortality rates in 1980. The response rate of elderly mortalities per 1,000 people over the age of 75 per year is observed to be proportional to (and, on these scales, double) the concentration of PM2.5. Does this justify

Table I. What Conclusions Do These Data Warrant About How Changing PM2.5 Would Change Elderly Mortality Rates?

Community	PM2.5 in 1980 ($\mu g/m^3$)	Elderly Mortality Rate in 1980 (per 1,000 People Over Age 75 per Year)
A	4	8
B	8	16
C	12	24

Table II. Should the Estimated PM2.5-Mortality C–R Relation Be the Same as for Table I?

Community	PM2.5 in 1980 ($\mu g/m^3$)	Income	Elderly Mortality Rate in 1980
A	4	100	8
B	8	60	16
C	12	20	24

an inference that reducing exposure concentration would reduce elderly mortality rates? Does it justify the stronger conclusion that every 10 $\mu g/m^3$ change in PM2.5 should be expected to produce a corresponding change of 20 mortalities per 1,000 elderly people per year? (Assume very large sample sizes, as the point of this example is not to quibble about the hypothetical numbers, but to scrutinize the logic of what may be validly inferred from such data.)

Before answering, consider Table II, which shows the same data as Table I, augmented by an additional column for average per household income per year in each community (in thousands of dollars). The income numbers are lower where the PM2.5 numbers are higher, so that higher PM2.5 is associated with lower income.

Given these data, an investigator who believes that income is irrelevant and that PM2.5 is a potentially important cause of elderly mortality might fit the following structural equation (causal) model to the data:

$$\textit{Average Elderly Mortality Rate} = 2 \times \textit{PM2.5 Concentration}\ (\textit{Model}\ 1)$$

Interpreting this causally, a change in the right-hand side variable, PM2.5 concentration, is predicted to cause an adjustment in the left-hand side (dependent) variable, elderly mortality rate, until equality is restored. Every 10 $\mu g/m^3$ increase in PM2.5 is then predicted to produce a corresponding increase of 20 mortalities per 1,000 elderly people per year. On the other hand, a different investigator, persuaded that income matters a lot but agnostic about the effects of ambient levels of pollution matter, might fit the alternative Model 2 to the same data.

$$\textit{Average Elderly Mortality Rate} = 35 - 0.5 \times \textit{PM2.5 Concentration} - 0.25 \times \textit{Income}\ (\textit{Model}\ 2)$$

Interpreted causally, Model 2 predicts that each 10 $\mu g/m^3$ increase in PM2.5 will cause a *reduction* of 5 mortalities per 1,000 elderly people per year. (Empirically, in a study of 27 U.S. communities, about a third of observed estimated C–R relations for PM2.5 and mortality were negative, three of them significantly so.[6] Toxicologically, relatively low levels of exposure to particulate matter can result in up-regulation of the production of endogenous protective antioxidants in the lung, although higher exposure concentrations overwhelm this protective effect with increased reactive oxygen and nitrogen species (ROS and RNS).[7] Thus negative coefficients may not be altogether far-fetched, notwithstanding the claim of Shin *et al.* that "any reported negative statistical estimates of the relationship between PM2.5 and mortality must be due solely to statistical error.")

Which C–R regression coefficient, 2 in Model 1 or –0.5 in Model 2, should be considered to characterize "the" C–R relation between PM2.5 and elderly mortality? In principle, the answer is not determined by the data: Models 1 and 2 both fit the data equally well. In practice, it is determined by which model the investigator chooses. (In practice, also, it is unlikely that different models will all explain the data perfectly, as in this example. But the same key point holds: different models that fit the data approximately equally well, and better than other models, often yield very different conclusions.) Thus, a study that cites Model 1's regression coefficient of 2 as evidence that reducing PM2.5 would reduce elderly mortality should not be accepted as credible, as it shows only that the investigator selected Model 1 instead of Model 2. It does not necessarily reveal how future real-world changes in PM2.5 would change real-world mortality rates. Similarly, concluding on the basis of Model 2 that reducing PM2.5 would increase elderly mortality would simply reflect a choice of a model that implies this conclusion instead of a different model that implies its opposite. Such ambiguous associations—that is, associations that depend on modeling choices, and that can easily

be reversed by varying the modeling assumptions—make the conclusions from association-based methods (including traditional C–R modeling) unreliable.

Real-world data frequently exhibit the key feature of this example: signs and magnitudes of the estimated C–R relation vary widely with modeling choices.[5] For example, Table III shows a Poisson regression model for number of daily mortalities among people 75 or older (a count variable) fit to the Los Angeles data set (available in the previously mentioned CAT package), using the generalized linear modeling ({glm}) package in R. To facilitate replication and reanalysis for interested readers who do not use R, the Appendix provides an Excel alternative to the R analyses. Table III shows the result of regressing daily mortality counts for people aged 75 or older, denoted by *AllCause75*, against same-day values of minimum and maximum daily temperature (*tmin* and *tmax*), maximum relative humidity (*MAXRH*), and the *month* and *year* for the observations. The regression coefficient for PM2.5 is significantly positive, $p = 0.00038$. If an alternative Poisson regression model is fit to the same data, with *month* treated as a discrete factor instead of as a continuous predictor (with possible values of January–December coded as 0–1 dummy variables by replacing *month* with *as.factor(month)* in the R model), then PM2.5 is no longer a significant predictor of elderly mortality at the conventional 0.05 significance level ($p = 0.06$ in a Poisson model, $p = 0.09$ in a quasi-Poisson model). If values of PM2.5 lagged by one, two, or three days are included as predictors, then the only significant C–R coefficient between PM2.5 and elderly mortality counts is negative ($p = 0.006$ if month is treated as continuous; $p = 0.03$ if month is treated as a discrete factor), at a lag of three days.

Thus, the finding of a significant positive C–R relation between PM2.5 and elderly mortality counts in Table III is in a very real sense *created* by modeling choices to "control" for the effects of month using a relatively inflexible (e.g., linear) model and to include only same-day values of variables as predictors. Different modeling choices remove or reverse the finding, consistent with the warning of Dominici *et al.*[5]

In practice, investigators often use more sophisticated models (e.g., splines with an investigator-specified stiffness and number of knots to account for the smoothed effects of seasons and trends; or varying combinations of lags for different predictors). But the fundamental methodological problem remains: any significant C–R associations found may simply reflect the particular modeling choices made. This is not a problem that can easily be overcome by appeals to expert judgments, for example, by inviting selected experts to opine about the causal interpretation of reported findings, insofar as the experts themselves do not know what other models would have shown. Sensitivity analyses for a selected model may reveal the sensitivity of its conclusions to variations in its assumptions, but without revealing whether taking very different models as a starting point would have led to very different conclusions. Model ensemble methods, especially with nonparametric models (such as the popular random forest algorithm, discussed later) provide a possible constructive solution to these challenges by examining the distribution of C–R estimates from hundreds or thousands of models, but have been criticized on the grounds that their results do not necessarily reflect the beliefs (or "substantive knowledge") of subject matter experts.[8]

2. ASSOCIATIONS BETWEEN LEVELS DO NOT PREDICT ASSOCIATIONS BETWEEN CHANGES

Suppose that all PM2.5 levels in Table II were to be cut in half between 1980 and 1990, as shown in Table IV. Given the data available in 1980, can the effects on elderly mortality of this reduction in PM2.5 be predicted? Specifically, what will the mortality rates for communities A–C in 1990 be, and how sure can we be?

This prediction challenge highlights the distinction between finding an associational C–R relation that *describes* the association between past levels of exposure concentrations and past levels of responses (as Models 1 and 2 do for the data in the top half of Table IV) and developing a valid causal C–R relation that *predicts* how changing the levels of exposure concentrations would change the responses, allowing the ? values in the bottom half of Table IV to be predicted correctly. Different models and methods are typically needed for these two different purposes. But the C–R literature to which Shin *et al.*,[1] Fann *et al.*,[2] Smith,[3] and Anenberg *et al.*[4] refer thoroughly conflates them.

Although it may be reasonable for policymakers to ask risk analysts how halving exposure would affect mortality rates, the unfortunate truth is that the data in the top half of Table IV do not permit valid answers to the question: they place no constraints on the possible values of the three unknown (in 1980) quantities in the lower right, that is, the 1990

Table III. A Poisson Regression Model for Daily Elderly Mortality Counts (AllCause75)

Dependent Variable: AllCause75				
Estimated Coefficients				
	Estimate	Std. Error	Z Value	Pr(>|z|)
(Intercept)	3.684524	4.126827	0.89	0.37195
PM2.5	0.000745	0.000210	3.55	0.00038 ***
tmin	–0.003820	0.000517	–7.38	1.6e-13 ***
tmax	–0.001776	0.000369	–4.81	1.5e-06 ***
MAXRH	–0.000961	0.000194	–4.96	7.0e-07 ***
year	0.000833	0.002056	0.41	0.68536
month	–0.009686	0.000668	–14.50	<2e-16 ***

Significance codes: 0 "***": 0.001 "**": 0.01 "*": 0.05 ".": 0.1 " "1.
Source: The numbers in Table III can be obtained by applying the following R commands applied to the LA data: fit ← glm(AllCause75 ~ PM2.5 + tmin + tmax + MAXRH + month + year, data = data.frame(PM2.5, tmin, tmax, MAXRH, month, year), family = poisson()): summary(fit).

Table IV. Effects on Mortality of Future Changes in Exposure Are Underdetermined by Past Exposure–Response Data

Community	Year	PM2.5 ($\mu g/m^3$)	Income	Elderly Mortality Rate in 1980
A	1980	4	100	8
B	1980	8	60	16
C	1980	12	20	24
A	1990	2	100	?
B	1990	4	60	?
C	1990	6	20	?

mortality rates in the three communities. For example, if the values of these three quantities (from top to bottom) turn out in 1990 to be 8, 16, and 24, the same as in 1980, then we will have learned that changes in PM2.5 appear to have had no impact, and that perhaps only income matters for predicting mortality rates. If instead they turn out to be 4, 8, and 12, then we might conclude that income has no discernible effect, and that halving exposure concentrations halves elderly mortality rates, in accordance with Model 1. If the numbers instead are eventually revealed to be 9, 18, 27, then we could conclude that Model 2 appears to have been right, proving that a positive association between C–R *levels* does not logically imply a positive association between future *changes* in their levels. (Decreases in PM2.5 of –2, –4, and –6 $\mu g/m^3$ in communities A, B, and C from 1980 to 1990 would correspond to increases in daily mortality of 1, 2, and 3 deaths per day, respectively, even though daily mortality is proportional to PM2.5 in 1980.) Other outcomes in which both income and pollution affect mortality in various ways can readily be envisioned. The key point is that the data available in 1980 provide no sound logical or statistical basis for predicting what the three values denoted by "?" in the lower right of Table IV will turn out to be in 1990. Specifically, it would be incorrect to assume that a C–R relation such as Model 1 or Model 2 estimated from data in the top half of Table IV allows prediction of how changes in PM2.5 would change elderly mortality rates in the bottom half of Table IV. (In the terminology of econometrics, this assumption confuses a "reduced-form" regression equation, which is what air pollution researchers typically work with, with a "structural" or causal equation, which is what policymakers need.) It is instead necessary to learn about how the world works—specifically, how changing one or more variables changes others that depend on them—rather than assuming that past associations can be used to predict future changes. Asking experts whether they think that historical associations are causal, or whether concentration is "a causal factor" (to an unspecified degree) for response, in the terminology of Fann *et al.*,[2] does

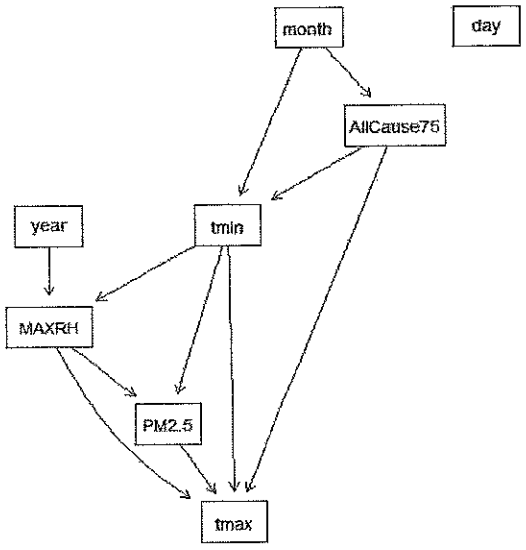


Fig. 1. A Bayesian network (BN) model for the LA data, showing statistical dependencies among variables, indicates that *AllCause75* does not depend directly on *PM2.5*.
Source: Fig. 1 was generated by running R package {*bnlearn*} on the LA data using the default hill-climbing (*hc*) method, as follows: library(bnlearn); library(Rgraphviz); bn ← hc(data.frame (year, month, AllCause75, PM2.5, tmin, tmax, MAXRH)); graphviz.plot(bn, shape = "rectangle").

not solve the problem that associational models do not in general reveal how future changes in concentrations will affect future changes in response.

3. TOWARD MORE RELIABLE AND PREDICTIVELY USEFUL CAUSAL C–R RELATIONS

The foregoing comments do not imply that models cannot or should not be used to predict (or at least to constrain predictions about) how changes in exposure concentrations would change responses. They only imply that quantifying C–R associations between past concentration and response levels, for example, using reduced-form regression models, and then extrapolating these C–R relations to predict future changes is not the way to do it. But many other risk analysis methods and models, from simulation to causal Bayesian networks, are available to help identify, test, and quantify the changes in responses caused directly and indirectly by changes in exposures.[9] For example, Fig. 1 shows the structure of a Bayesian network (BN) model learned directly from the LA air basin data for Table III via a data-mining algorithm that searches through a set of many possible models to find one that best explains the data. All variables in the LA data set are included, although *day* of the month turns out not to significantly affect the frequency distribution of any other variable in the data set; thus, it appears as an isolated node in Fig. 1.

In Fig. 1, the nodes (rectangles) represent the variables, that is, columns in the underlying data set. The arrows signify "is not statistically independent of" (or, more colloquially, "is informative about"). An arrow from X to Y implies that the conditional frequency distribution of Y differs significantly for at least some different values of X (and for at least some configuration of values of other variables pointing into Y, if any); otherwise, they remain unlinked. Arrows between variables can typically point in either direction, by Bayes's Rule. (Technically, the arrow directions in Fig. 1 indicate one of many ways to compute the joint distribution of the variables from the marginal distributions of the input variables—those with only outward-pointing arrows [here, *year* and *month*]—and the conditional probability tables for all other variables, specifying the conditional probability for the value of each variable for each combination of values of the variables that point into it. Continuous variables are automatically binned into deciles to permit these conditional probability calculations. For more details on Bayesian networks and structure learning algorithms, see https://cran.r-project.org/web/packages/bnlearn/bnlearn.pdf and the references therein.)

In Fig. 1, *AllCause75* does not detectably depend on *PM2.5* once the daily minimum and maximum temperatures *tmin* and *tmax* and *month* are known; more accurately, *PM2.5* provides no further information that improves ability to predict *AllCause75*. In this sense, the observed data provide no evidence that *PM2.5* directly affects *AllCause75*. Interventions that change the daily temperatures experienced by elderly people might affect both *PM2.5* and *AllCause75*, but interventions that change only *PM2.5* and not daily temperature should not be expected to change elderly mortality rates, even if regression models show a statistically significant C–R association between them, as in Table III. Relations between conditional independence, interventions, and prediction of causal effects are discussed further in the causal modeling and causal graph literature and supporting software packages (see e.g., https://cran.r-project.org/web/packages/causaleffect/causaleffect.pdf and references therein, and Cox[9]).

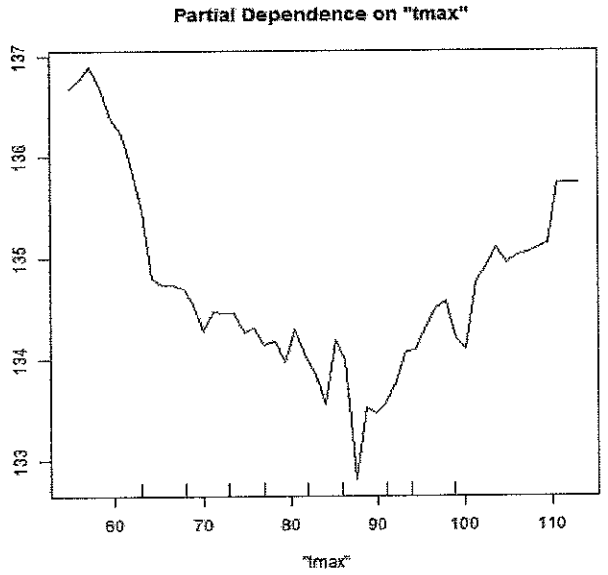


Fig. 2. A partial dependence plot for *AllCause75* versus *tmax*, showing that daily elderly mortality counts are predicted to be smallest for days with maximum daily temperature of about 86 °C. *Source*: Fig. 2 is generated by applying the *partialPlot*() function in the {*randomForest*} R package to the LA data, as follows: library(randomForest); data ← data.frame(year, month, day, AllCause75, PM2.5, tmin, tmax, MAXRH); partialPlot (randomForest(data, AllCause75), pred.data = data, x.var = "tmax").

Although the algorithms that search through alternative Bayesian network models to identify those that best explain the data can be complex, and the interpretation of network diagrams such as Fig. 1 can be subtle, it is easy to understand the main insights from such automated analyses: that temperature and humidity are associated with (more precisely, informative about) PM2.5 and that if mortality is driven primarily by these meteorological variables and not PM2.5 levels, then PM2.5 will be associated with mortality rates, but weather variations, rather than fluctuations in PM2.5, contribute causally to fluctuations in mortality.

Beginning with such data-driven findings about the qualitative structure of dependencies among variables, causal modelers and risk analysts can then quantify the relations between outcomes of interest and factors that might affect them (e.g., those to which they are linked by arrows in a BN) without having to commit to any specific model or small set of models. For example, Fig. 2 shows a partial dependence plot for how elderly mortality (*AllCause75*) varies with maximum daily temperature (*tmax*) when only *tmax* is assigned alternative (counterfactual) values and all other variables have their actual values in the data set.

(Further background on random forest algorithms, which are among the most successful machine-learning techniques, and partial dependence plots is provided at https://cran.r-project.org/web/packages/randomForest/randomForest.pdf. The partial dependence plot is similar to an added variables plot for a regression model, but uses a random forest model ensemble of many classification and regression trees instead of a single regression model to predict the value of a dependent variable for different values of an independent variable. For the LA data set, the random forest model ensemble explains 41% of the variance in *AllCause75*, as contrasted with about 32% explained by the best single linear regression model.)

The nonparametric model ensemble that generated Fig. 2 easily detects and quantifies the nonlinear dependence of elderly mortality counts on *tmax*. It could as easily do the same for PM2.5 if elderly mortality counts depended on PM2.5 levels. It is based on an ensemble of several hundred automatically generated nonparametric models (about 90 suffice to saturate the predictive accuracy of the ensemble), and thus avoids the need to make model selection and specification choices that might otherwise undermine the reliability of the results. Ensembles of BN models can be developed and used similarly.

The challenge that future changes in responses caused by changes in concentrations are underdetermined by data (Table IV) can be substantially met with the help of causal graph algorithms developed to check sufficient conditions for transporting observed conditional probability relations across contexts (e.g., the {*causaleffect*} package in R) and time series methods developed to check the stationarity of relations among different time series (e.g., the {*changepoint*} and {*ecp*} packages in R). Finally, BNs and related graphical models offer a more nuanced description of causation than "judgments about the likelihood that PM2.5 is a causal factor in mortality"[2] by making explicit both direct and indirect paths between exposure concentration and response variables and by allowing the *fraction* of a C–R association that is due to a direct link between them (if any) to be estimated distinctly from the fractions due to other pathways. These methods allow genuine prediction of how changes in exposures will change effects once genuine, stable causal relations or laws have been identified and validated. They can augment older regression-based models to

substantially address the objections raised in the examples of Tables I–IV.

ACKNOWLEDGMENTS

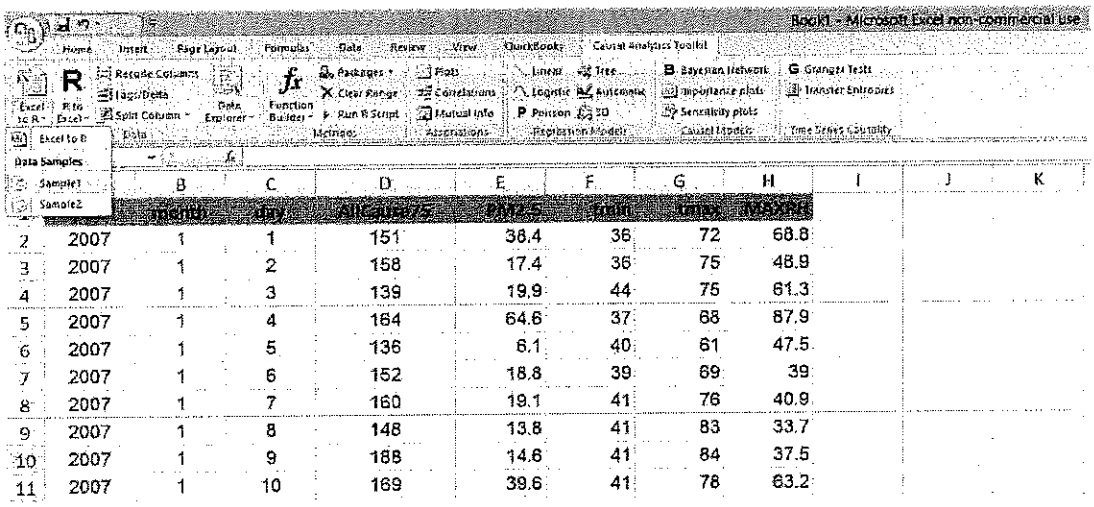
I thank Area Editor Warner North for a close reading and very valuable suggestions that helped simplify and clarify my exposition of several technical points. Warner also suggested introducing an appendix and references for more detailed technical exposition. The final exposition has been improved by his insightful advice.

APPENDIX: CAUSAL ANALYTICS TOOLKIT (CAT) FOR EXCEL USERS

A free Causal Analytics Toolkit (CAT) package may be downloaded from https://regulatorystudies.columbian.gwu.edu/causal-analytics-toolkit-cat to give readers access to the Los Angeles data set provided by Dr. Stanley Young and to enable independent replication, new analyses, or extensions of the analyses summarized in Table III and Figs. 1 and 2. CAT is an Excel add-in that provides relatively simple commands and a point-and-click interface for doing advanced analytics from Excel using R packages, even if the user does not know R. A user guide that describes how to download and install CAT is available here: http://cox-associates.com/CAT/UserGuide.pdf.
Once CAT has been downloaded and installed, it can be enabled as an Excel add-in (select File > Options > Add-Ins > Go and check Causal Analytics Toolkit, then click OK to add it to the Excel toolbar). To install the LA data used in Table III and Figs. 1 and 2, open a new Excel workbook and click on the "Excel to R" button at the far left of the CAT ribbon; then select "Sample1" to download the data. The newly created data set should look as shown in the screen shot (with 1,401 records) on an automatically-created sheet named "Data."

Select/highlight the data (columns A–H) and click on the Excel-to-R button at the far left of the CAT ribbon to import the data to R. (Click "Yes" to accept the default of creating an R data frame called "Data" corresponding to the data on the Data worksheet.) The data are now ready for analyses using CAT.

To replicate the Poisson regression model in Table III, select the dependent variable, *AllCause75* (i.e., click on column D to highlight it). Then use Ctrl-click to select the independent variables. (Multiple adjacent columns E–H can be selected by swiping across them while holding down the Ctrl key. Then, go to column B and use Ctrl-click to select *month* for inclusion in the model.) Click on *P Poisson* in the Regression Models area of the CAT ribbon to generate a new sheet with the results of the Poisson regression model; this is how Table III was produced. For users who would rather type than point and click, entering R: CAT_poisson(AllCause75, PM2.5, tmin, tmax, MAXRH) in any cell will produce the same Poisson regression model output immediately below that cell. (Once CAT is installed, entering "R:" into any Excel cell makes it behave like an R command console, ready to receive and process R commands.) Returning to the Data sheet with the columns already highlighted, and using Ctrl-Click to add *year* and *day* (or simply reselecting all columns A–H), and



	B	C	D	E	F	G	H	I	J	K
	month	day	AllCause75	PM2.5	tmin	tmax	MAXRH			
2	2007	1	1	151	36.4	36	72	68.8		
3	2007	1	2	158	17.4	36	75	48.9		
4	2007	1	3	139	19.9	44	75	61.3		
5	2007	1	4	164	64.6	37	68	87.9		
6	2007	1	5	136	6.1	40	61	47.5		
7	2007	1	6	152	18.8	39	69	39		
8	2007	1	7	160	19.1	41	76	40.9		
9	2007	1	8	148	13.8	41	83	33.7		
10	2007	1	9	188	14.6	41	84	37.5		
11	2007	1	10	169	39.6	41	78	63.2		

then clicking on *B Bayesian network* in the Causal Models area of the CAT ribbon will run the {*bnlearn*} package and generate Fig. 1 (using ellipses rather than rectangles as the default shape for nodes). Fig. 2 can be generated either by selecting the columns in the order *AllCause75*, *tmax*, and then the remaining variables in any order and clicking on *Sensitivity Plots* in the CAT ribbon; or by entering the following R commands successively into any three Excel cells.

R: library(randomForest)
R: data ← data.frame(year, month, day, AllCause75, PM2.5, tmin, tmax, MAXRH)
G: partialPlot(randomForest(data, AllCause 75), pred.data = data, x.var = "tmax")
CAT uses the "G:" prefix to direct graphics output to the spreadsheet; if "R:" is used instead, then it will appear in a separate window.

The preceding instructions should suffice to allow replication of Table III and Figs. 1 and 2. For new analyses, the following CAT functions are useful. (Many of these can also be accessed through the point-and-click ribbon features, as just described, and also through a Function Builder interface that allows users to select the name of the CAT function and then its arguments from drop-down lists.)

- *CAT_describe*(X) generates summary statistics and plots (e.g., frequency distribution histograms) for selected variable X. (The "Data Explorer" feature of CAT allows such results to be viewed simply by passing the cursor over a column with data in it.)
- *CAT_correlations*(...) and *CAT_associations* (...) will display graphs and tables showing various types of correlations and associations among the variables (e.g., AllCause75, tmin, tmax), generically indicated by "..." specified by the user.
- *CAT_regression*(Y, X, ...) will automatically select an appropriate regression model (linear if all variables are continuous, and also by default; logistic if Y is 0–1, and Poisson if Y is a count variable), fit it to the data, and generate various tables and plots with Y as the dependent variable and X, ... (the names of one or more other variables) as independent variables.
- *CAT_tree*(Y, X, ...) will generate a classification and regression tree for dependent variable Y (typically a health effect) and independent variables X, ... (typically an exposure variable and other covariates).
- *CAT_bnLearn*(...) will display the structure of a Bayesian network, given a list of comma-separated variable names (e.g., AllCause75, tmin, tmax), generically indicated by "...," to be included in the model.
- *CAT_show3d(Y, X, Z)* generates a 3D scatter plot of response Y against variables X and Z and fits a smooth surface for the expected value of Y given X and Z. (Multiple surfaces for different values of a discrete fourth variable, W, can be generated with CAT_show3d(Y, X, Z, W).)
- *CAT_grangerTests*(X, Y) will assess whether X is a Granger-cause of Y over a horizon (default is 7 time steps) that the user can specify, if X and Y are both time series variables.

Any of these functions can be invoked by typing the prefix R: followed by the CAT function in any cell of an Excel spreadsheet, once the CAT add-in is installed.

CAT is intended to make advanced analytics readily available to Excel users, and it can be used to install new R packages as desired and access them through the CAT interfaces (specifically Function Builder, which works with all R packages and functions). New functions may be added to the CAT ribbon over time.

REFERENCES

1. Shin HH, Cohen AJ, Pope CA 3rd, Ezzati M, Lim SS, Hubbell BJ, Burnett RT. Meta-analysis methods to estimate the shape and uncertainty in the association between long-term exposure to ambient fine particulate matter and cause specific mortality over the global concentration range. Risk Analysis. 2016(9) (this issue).
2. Fann N, Gilmore EA, Walker K. Characterizing the long-term PM concentration-response function: Comparing the strengths and weaknesses of research synthesis approaches. Risk Analysis, 2015; 36(9):1693–1707.
3. Smith AE. Inconsistencies in risk analyses for ambient air pollutant regulations. Risk Analysis, 2016; 36(9):1737–1744.
4. Anenberg SC, Belova A, Brandt J, et al. Survey of air pollution health risk assessment tools. Risk Analysis. 2016; 36(9):1718–1736.
5. Dominici F, Greenstone M, Sunstein CR. Science and regulation. Particulate matter matters. Science. 2014; 344(6181):257–259.
6. Franklin M, Zeka A, Schwartz J. Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. Journal of Exposure Science and Environmental Epidemiology. 2007; 17(3):279–287.

7. CDC (Centers for Disease Control and Prevention (US)); National Center for Chronic Disease Prevention and Health Promotion (US); Office on Smoking and Health (US). How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention (US), 2010. 7, Pulmonary Diseases. Available at: http://www.ncbi.nlm.nih.gov/books/NBK53021/. Accessed August 1, 2016.
8. Thomas DC, Jerrett M, Kuenzli N, Louis TA, Dominici F, Zeger S, Schwartz J, Burnett RT, Krewski D, Bates D. Bayesian model averaging in time-series studies of air pollution and mortality. Journal of Toxicology and Environmental Health. 2007; 70(3–4):311–315.
9. Cox LA Jr. Quantifying and reducing uncertainty about causality in improving public health and safety. In Graham R, Higdon D, Owhadi H (eds). Handbook of Uncertainty Quantification. New York: Springer, 2016.