Review and Critique of the U. S Environmental Protection Agency Second External Review Drafts of the "Health Risk and Exposure Assessment for Ozone" and the "Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards"

By
Jon M. Heuss
George T. Wolff
Dennis F. Kahlbaum
Air Improvement Resource, Inc.

Prepared for
The Alliance of Automobile Manufacturers

March 24, 2014

Executive Summary

Air Improvement Resource, Inc. (AIR) reviewed the second draft Health Risk and Exposure Assessment for Ozone (HREA) and the second draft Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards (PA). AIR focused on the portions of the documents that are important to providing the Administrator with the most relevant science with which to judge the health effects of ozone and establish a primary ozone air quality standard that will protect the public health.

AIR comments address the background of ozone uncontrollable through reduction in U.S. man-made emissions, the human clinical studies of ozone effects and their interpretation in terms of public health, and the epidemiological studies of associations of ozone with health endpoints and their interpretation in terms of public health.

Comments on Background Ozone

There are two fundamental issues involving background ozone. The first is what EPA uses for background ozone which depends on how it is defined and how it is estimated. The second is how they use background in their policy and risk assessments which is then used to inform policy decisions that need to be made concerning the form, averaging time and level of the National Ambient Air Quality Standard (NAAQS).

After a long evolutionary process EPA has finally chosen to use U.S. background (USB) O_3, defined as the O_3 that exists in the U.S. in the absence of anthropogenic U.S. emissions, as the policy relevant background. While we agree with the choice of USB,

we believe that that their estimation methods are flawed and result in estimates that are biased low.

EPA calculates USB using two methods: the CMAQ modeling system with zeroed out U.S. anthropogenic emissions and the CAMx modeling system with the Anthropogenic Precursor Culpability Assessment (APCA) tool. The first method creates a hypothetical atmosphere that give unrealistic results due to the nonlinear chemistry. We have been critical of this approach in our previous comments as well.

Unfortunately, the APCA tool was not designed to determine how much of the modeled ozone has resulted from background sources, but, rather, as its name implies, to attribute maximum culpability to controllable anthropogenic precursors. As a result, the USB estimates generated by the APCA tool are negatively biased. The best way to illustrate this is with examples.

Example 1: A biogenic VOC molecule interacts with an anthropogenic NO_x molecule to form an O_3 molecule. CAMx/APCA counts the resulting O_3 as an anthropogenic O_3 molecule and not USB even though it would not exist without the contribution of the biogenic VOC. Similarly a naturally emitted NO_x molecule that reacts with an anthropogenic VOC to form O_3 would also be counted as an anthropogenic O_3 molecule.

Example 2: An O_3 molecule entering the U.S. from Canada is initially counted as boundary condition O_3 which is included in USB. However, should that molecule encounter an NO molecule emitted from an anthropogenic source and react to form NO_2, which is subsequently photolyzed and results in the formation of an ozone molecule, that ozone molecule is counted by APCA as anthropogenic even though it would not have existed if not for the parent O_3 molecule that traveled from Canada.

Example 3: An O_3 molecule enters the U.S. and encounters an anthropogenic VOC (olefin) which react to form an aldehyde which is subsequently photolyzed to form an ·OH radical. The ·OH radical is capable of participating in a chain reaction with NO_x and VOCs to produce many O_3 molecules. APCA would count all of the O_3 formed in this manner as anthropogenic even though they would not have existed if not for the imported initial O_3 molecule.

There are numerous other examples that could be used. For example, an O_3 molecule formed naturally in the stratosphere and transported to the troposphere would be subjected to the same accounting procedure as the ones that travel across the border into the U.S.

APCA is designed to identify the controllable anthropogenic emissions that if reduced would result in lower ozone. In the case of USB, it identifies the additional reductions that would be needed to offset the USB impacts and not the contribution of USB to observed O_3. In reality, the impact of imported O_3 or natural O_3 can be propagated throughout the US by subsequent generations of reaction products that would not exist if not for the initial O_3 molecules that were transported into the U.S. or formed naturally. In

the APCA accounting procedure, the impact of these initial molecules is terminated as soon as they react with a molecule of anthropogenic origin. Thus, the CAMx/APCA modeling system underestimates the impact of USB on ambient O_3 concentrations.

As we recommended in our earlier comments, there are two ways to obtain more realistic estimates of the impacts of USB on ambient ozone concentration in the U.S. The first way is to run CAMx (we prefer CAMx over CMAQ because of the better performance of CAMx) with all emissions and then with the boundary conditions and natural sources zeroed out. The difference between the two scenarios provides the impact of USB on U.S. O_3 concentrations. The second way is to run CAMx with the APCA scheme modified to keep track of all the sources of odd oxygen atoms and distinguish between those that originated from the reactions involving natural emissions or imported O_3 or precursors from outside the U.S. and those formed from U.S. anthropogenic emissions alone.

In spite of the fact that we feel that the USB estimates made in the Policy Assessment are biased low, EPA's estimates still show that USB is a major component of the O_3 observed across the country. For 12 urban areas EPA uses as case study areas, the average fraction of the MDA8 O_3 due to USB range from 0.43 in Atlanta to 0.69 in Denver. The average of all the cities is about 0.6. This is a non-trivial contribution. On average, most of the MDA8 O_3 in most of the cities is from USB.

In the previous O_3 NAAQS review, EPA estimated O_3 exposure risks down to background O_3. In the present review, EPA has decided to calculate risks that are independent of the choice of background. In the second draft HREA, risks are estimated down to a concentration of zero ppb. By doing this, EPA has inflated the risk estimates. For the reasons discussed in the health effects section of these comments, EPA health risks are not realistic.

Because USB contributes significantly to MDA8 O_3 and because EPA does not exclude USB levels in their risk assessments, additional anthropogenic emission reductions will be needed to offset the impact of USB. As a result, extreme additional emission reductions will be required to achieve the alternative NAAQS being considered by EPA. EPA provides "ball park" estimates for the emission reductions that will be required in 15 cities to meet the various alternative standards that are being considered. These are shown in Table ES-1. To achieve a 70 ppb NAAQS, NO_x reductions of 27 - 89% are needed and to achieve a 60 ppb NAAQS, NO_x reductions of 62 - 93% are needed.

Urban Area	Years	Standard Level*			
		75 ppb	70 ppb	65 ppb	60 ppb
Atlanta	2006-2008	50%	58%	64%	71%
	2008-2010	23%	43%	54%	62%
Baltimore	2006-2008	46%	54%	61%	69%
	2008-2010	44%	52%	60%	67%
Boston	2006-2008	40%	49%	61%	70%
	2008-2010	13%	40%	53%	65%
Chicago	2006-2008	19%	52%	66%	80%
	2008-2010	N/A	27%	55%	70%
Cleveland	2006-2008	48%	61%	73%	88%
	2008-2010	50%	64%	77%	88%
Dallas	2006-2008	50%	57%	65%	72%
	2008-2010	50%	58%	64%	71%
Denver	2006-2008	51%	65%	78%	87%
	2008-2010	15%	46%	64%	87%
Detroit	2006-2008	59%	69%	76%	84%
	2008-2010	N/A	54%	66%	78%
Houston	2006-2008	62%	68%	74%	82%
	2008-2010	42%	53%	63%	75%
Los Angeles	2006-2008	87.1%	89.3%	91.2%	93.2%
	2008-2010	87%	89%	91%	93%
New York	2006-2008	64%	74%	92%	N/A
	2008-2010	52%	67%	89%	N/A
Philadelphia	2006-2008	54%	61%	68%	74%
	2008-2010	42%	52%	61%	68%
Sacramento	2006-2008	63%	70%	76%	84%[5]
	2008-2010	64%	71%	77%	84%
Saint Louis	2006-2008	45%	56%	66%	75%
	2008-2010	10%	34%	50%	63%
Washington D.C.	2006-2008	53%	60%	67%	74%
	2008-2010	31%	50%	60%	71%

* N/A values for the 75 ppb standard level mean that a particular urban area did not have any design values above 75 for that 3-year period so no controls were needed. N/A values for the 60 ppb standard level mean that this adjustment methodology was not able to bring design values down to 60 for that particular city and 3-year period.

Table ES-1: Percent emission reductions used for each urban area to achieve each alternative NAAQS. Percentages in Chicago and Denver represent reductions in both anthropogenic NO_x and VOC. Percentages in all other cities represent reductions in NO_x only.

Comments on the HREA

The controlled human exposure studies provide a strong body of information on the dose-response of effects of 1- to 3-hour and 6- to 8-hour exposures to ozone. The first effects - transient, reversible FEV1 decrements – are evident after exposures to 80 ppb for 6 to 8 hours when the subjects are exercising at a rate that would be considered very strenuous when carried out for an eight-hour period. The HREA uses the same exposure modeling methodology used in the prior review to calculate the number of exposures and number of FEV1 decrements above various benchmark concentrations with exercise. AIR

demonstrates how the EPA exposure model over-estimates the number of exposures with high ventilation rates in the population.

Nevertheless, using EPA's own model, the fraction of person-days with children experiencing FEV1 decrements under current air quality is extremely small. For example for Denver in 2006, when the design value was 90 ppb, the portion of persons-days for children with FEV1 decrements >15 % is estimated as 0.00023 or 0.023 %. Thus using the HREA methodology, current air quality is very protective of public health. Attainment of the current standard would reduce these already extremely small risks. In addition, physiological responses of this nature from single exposures have not been considered adverse in prior reviews. To provide a more complete perspective on the public health impact of the current and alternative standards, the final HREA should correct the exposure model for the biases identified by AIR and present the estimates for both persons with one or more occurrences in an ozone season and person-days of occurrence.

The HREA uses two approaches to estimate FEV1 decrements. Estimates using the first approach are similar to the estimates made in the prior review. Estimates of decrements using the second approach, the MSS model, are somewhat higher. However, due to its high variability, the MSS model predicts some individual decrements at very low exposures and low exercise levels where the group mean decrements are extremely small. It is not clear whether these are actual effects due to ozone or whether they are related to the noise in the underlying data. With both approaches, the portion of person-days with FEV1 decrements is extremely small at the current standard.

The epidemiological or observational studies of the association of ozone with various health endpoints continue to be difficult to interpret. Based on AIR's review, EPA made choices as to which associations to include in the core analyses, how to model the concentration-response functions, and as to the way the analyses are presented in the HREA that dramatically overstate the magnitude and certainty of ozone health risks.

For example, the HREA uses selected results from the Smith et al. (2009) analysis. However, the authors of that study concluded:

> …the heterogeneity and sensitivity of ozone effect estimates to a variety of covariates leaves open the issue of whether or not ozone is causally related to mortality. Consequently, the question arises whether any particular ozone-mortality effect estimate can reliably be used to predict mortality reductions that would ensue from specific ozone reductions.

The authors also cautioned that it is possible that the appearance of an association at low ozone levels may be due to the effect of co-pollutants, or an artifact caused by differences between personal and ambient exposure.

The HREA estimates risk based on a mix of positive ozone associations from single-city studies and Bayes-adjusted city-specific effect estimates from selected multi-city studies. Even so, the mortality risk in most of the 12 cities evaluated is not statistically significant. AIR demonstrates that if the unadjusted city-specific effects are used, the risks vary from positive to negative, covering a range that is biologically impossible. AIR demonstrates that model selection uncertainty is extremely large compared to the EPA estimates of risk and that there is a temporal and spatial pattern to the data that is not consistent with ozone causality.

By assuming ozone mortality extends down to zero ozone and by using selected ozone-mortality associations from the literature, the HREA calculates a substantial burden of mortality even when man-made emissions are taken away. In fact, a whole chapter of the HREA, Chapter 8, is devoted to the exercise of estimating a national mortality burden from ozone. However, the full pattern of associations in the literature is not consistent with ozone causing either acute or chronic mortality, the shape of the concentration-response is not known, and epidemiology studies cannot be used to identify a threshold because of exposure uncertainty. Consequently, EPA's extrapolations of risk at low ozone concentrations in the HREA are not justified.

By exploring the full range of spatial and temporal differences in association together with model selection uncertainty in the final HREA, the limitations of the epidemiologic risk assessment will become apparent. Given the variability and uncertainty in the observational studies, AIR recommends that they not be used to set regulatory standards.

Comments on the PA

The draft PA concludes that the health evidence and exposure/risk information call into question the adequacy of the public health protection provided by the current standard, that it is appropriate in this review to consider alternative standards that would increase public health protection, and that it is not appropriate to consider alternative standards with levels higher than the current standard.

AIR reviewed the draft PA as it relates to the primary NAAQS and concludes that it (1) overstates the nature and magnitude of ozone health effects and perceived risk to public health from current ozone levels, and (2) strains to make the case for inadequacy of the current ozone standard. The revisions that are necessary in the final HREA will have a major effect on the final PA and on the interpretation of the human clinical and observational data as well as on estimates of the risk to public health from the current ozone standard. Therefore, it is premature to make a judgment on the adequacy of the current standard before the needed revisions are made.

Chapter 3 of the PA, which summarizes the health evidence, overstates the consistency and coherence of the observational evidence. With regard to hospital admissions and mortality, the overall results of a large multi-continent Health Effects Institute (HEI) study do not support EPA’s claims of causal or likely relationships between ozone and these endpoints. In particular with regard to respiratory mortality, EPA makes claims for

consistent effects that are contradicted by the views of the original investigators and the HEI Review Committee. In addition, the issues of heterogeneity due to stochastic variability, model selection uncertainty, confounding, and publication bias are ignored or downplayed in the Chapter. The final PA should address all these issues in the interpretation of the observational studies and their integration with the full range of ozone effects studies.

The discussion of adequacy also needs to consider that the kind of effects identified in the most recent controlled human studies are mild, transient decrements in the performance of lung function tests generally unaccompanied by symptoms. They only occur near the current standard if the subject is exposed and exercising over an extended period of time at a rate that, when sustained for a long period, is at the very high end of real-world situations. Based on the EPA's estimates of the number of person-days of exposure above EPA's benchmarks with an even lower level of exercise, the fraction of person-days experiencing such effects is extremely low. These are rare occurrences at current ozone levels and will be even rarer occurrences when the current standard is attained. Isolated incidences of such effects have not been considered adverse in prior reviews. Thus, based on the controlled exposure studies, the current standard is highly protective of public health.

Another issue that needs to be fully vetted in the PA is that the existence of a substantial threshold for the first physiological effects in controlled studies is not consistent with EPA's assumption that the more severe effects suggested by some epidemiological associations have no threshold. Such an assumption is not consistent with either the general principles of toxicology or the specific findings of controlled ozone exposure studies. The final PA should address the issue of dose plausibility in detail.

AIR is concerned that the preliminary PA conclusion regarding adequacy relies on CASAC's previous advice regarding the level of the standard and does not consider the new information that (1) background ozone is much closer to the current standard than thought during the last review, (2) there is clear evidence for a threshold in the first physiological effects of ozone, (3) the risk based on person-days of exposure that might cause FEV1 decrements is extremely low at the current standard, and (4) the uncertainty as to whether ozone is causing hospital admissions or mortality is much larger than thought in the previous review.

Finally, the alternatives lower than the current standard that are evaluated in the PA are close to and may be exceeded by background ozone, as acknowledged in the PA. There is precedent for considering background in ozone NAAQS decisions. The 80 ppb 1-hour photochemical oxidant standard was revised to a 1-hour ozone standard of 120 ppb in 1979, in part, because there was evidence that it was too close to background. In the 1997 review, an 8-hour standard of 70 ppb was viewed as being too close to peak background. Therefore, the range of alternatives for the Administrator to consider should include the current standard.

Review and Critique of the U. S Environmental Protection Agency Second External Review Drafts of the "Health Risk and Exposure Assessment for Ozone" and the "Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards"

By
Jon M. Heuss
George T. Wolff
Dennis F. Kahlbaum
Air Improvement Resource, Inc.

Prepared for
The Alliance of Automobile Manufacturers

March 17, 2014 DRAFT

Introduction

The U. S Environmental Protection Agency (EPA) is in the process of reviewing the National Ambient Air Quality Standards (NAAQS) for ozone (O_3) with the issuance of the second external review drafts of the Health Risk and Exposure Assessment for Ozone (HREA)[1] and the Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards (PA).[2] Air Improvement Resource, Inc. (AIR) reviewed the two draft documents with a focus on the portions of the REA and PA that are important to providing the Administrator with the most relevant science with which to judge the health effects of ozone and establish a primary ozone standard which will protect the public health with an adequate margin of safety. AIR and the Alliance of Automobile Manufacturers (Alliance) participated in the previous review of the ozone standard that resulted in the 8-hour standard being set at 75 ppb (0.075 ppm).[3] AIR and the Alliance also participated in the re-consideration of the ozone standard that was initiated by Administrator Jackson in January 2010.[4] Finally, AIR and the Alliance provided public comments on the first, second, and third draft Integrated Science Assessments and on the

[1] U. S. Environmental Protection Agency, *Health Risk and Exposure Assessment for Ozone, Second External Review Draft,* EPA-452/P-14-004a, February 2014.
[2] U. S. Environmental Protection Agency, *Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards, Second External Review Draft*, EPA–452/P–14–002, January 2014.
[3] Comments of the Alliance of Automobile Manufacturers on EPA's Proposal to Revise National Ambient Air Quality Standards for Ozone, 72 Fed. Reg. 37,818 (July 11, 2007), dated Oct. 9, 2007.
[4] Comments of the Alliance of Automobile Manufacturers on EPA's Proposal to Revise National Ambient Air Quality Standards for Ozone, 75 Fed. Reg. 2992 (Jan. 19, 2010), dated Mar, 22, 2010.

first draft HREA and PA documents.[5]

The following comments focus on the background of ozone from non-U. S. sources and the way that background is considered in the review, on the human clinical studies of ozone effects and their interpretation in terms of public health, and on the epidemiological studies of associations of ozone with health endpoints and their interpretation in terms of public health.

Although the total body of scientific studies that inform the decision as to ozone air quality standards that will protect public health and welfare has not changed dramatically since the last review, there are three major changes in the science and one policy decision by the Agency that can substantially alter the outcome of the current review.

The first is the understanding of background ozone and its impact on the distribution of ozone concentrations. The choice of background ozone (the ozone that cannot be reduced through control of U. S, man-made emissions) is particularly important since it affects the risk estimates that the Agency will use later in the NAAQS review process and provides a limit to how stringent a standard can be and still be achieved throughout the U. S. As detailed in previous submissions (Alliance October 9, 2007 and March 22, 2010 comments), the Alliance has been concerned that EPA underestimated the relevant background in the prior review. As now acknowledged in the final ISA, this is the case and there is now substantial new modeling and other information that supports the Alliance view.

The second major change is that EPA now evaluates the distribution of ozone concentrations upon attaining alternative standards with photochemical modeling. This is a major improvement over the previous rollback technique because it includes consideration of the complex non-linear chemistry involved in ozone formation. As an added bonus, the modeling provides us with estimates of the emission control needed to meet the current standard as well as alternative standards. This information can be used to assess how close alternative standards are to peak background levels.

The third major change is that, as more epidemiological or observational studies of the association of ozone with various health endpoints are published, the fundamental weaknesses of this body of information have become more apparent. A great deal of stochastic variability, uncertainty due to model selection issues, potential confounding by other pollutants, and publication bias bedevil the interpretation of these studies. Public comments from AIR and from other scientists have detailed these concerns and inconsistencies.[6] However, the draft HREA and PA continue to gloss over the issues that

[5] G. T. Wolff, J. M. Heuss, and D. F. Kahlbaum, Review and Critique of the U. S Environmental Protection Agency First External Review Drafts of the "Health Risk and Exposure Assessment for Ozone" and the "Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards," Air Improvement Resource, Inc. report prepared for The Alliance of Automobile Manufacturers, October 2012.

[6] J. M. Heuss and George T. Wolff, Review and Critique of the U. S. Environmental Protection Agency's First External Review Draft of the "Integrated Science Assessment for Ozone and Related Photochemical Oxidants," Air Improvement Resource, Inc. Report, Prepared for The Alliance of Automobile Manufacturers, May 2011; C. R. Long, et al. "Comments on U.S. EPA's Causality Determinations for

have been raised in public comments and fail to address the uncertainty and inconsistencies that are present in the epidemiologic data. Instead, the Agency uses its preferred epidemiologic associations to assess the risk. As a result, the draft HREA and PA overstate the consistency and weight of evidence for ozone effects from observational studies.

In addition, there is a major disconnect between the risk assessment based on epidemiological studies and that based on human clinical studies. The human clinical studies clearly demonstrate that the first ozone effects that are mild and transient occur with a threshold of ozone dose due to the protective effects of antioxidants in the epithelial lining fluid. Only at concentrations above the current standard and with vigorous exercise does the dose approach effects that may be considered adverse. Yet the HREA assumes that ozone causes premature mortality down to zero ozone levels. The HREA and PA need to fully discuss this disconnect and discrepancy.

The human clinical studies of ozone are particularly important since these data provide a strong and consistent body of information on the dose-response of effects of 1- to 3- hour and 8-hour exposures to ozone. Although there are now more studies of 6- to 8-hour exposures to low ozone concentrations while exercising heavily, EPA's estimate of the dose-response curve at low concentrations has not changed appreciably. In addition, there is substantial information that the first effects (FEV_1 decrements, neutrophilic inflammation, and respiratory symptoms) all exhibit threshold behavior. The most important issue or question with regard to these data is how to translate the results into human risk as people go about their daily life. The HREA includes probabilistic modeling of ozone exposures that attempts to answer this question. As documented in the following, the draft HREA and PA substantially overestimate the risk from the effects identified in the clinical studies. The factors that lead to the overestimation of risk have been brought to the attention of the Agency several times in recent years, and are acknowledged in the body of the current draft HREA, but are ignored as the results of the HREA are summarized and then used in the PA.

The influential policy decision that EPA has made is to estimate risk from total ozone rather than the ozone controllable through reduction or even total elimination of U.S. man-made precursor emissions. This is a step backwards as discussed in detail below and obfuscates the impact on public health and welfare of choosing alternative standards.

Based on the risk assessment using the clinical studies, the current primary ozone standard is highly protective of public health. The risk assessment using even EPA's favored epidemiological associations and assumptions shows that the risk of mortality effects is small and highly uncertain. When the full range of associations in the literature

---

Short-term and Long-term Ozone Exposures and Mortality in the Integrated Science Assessment for Ozone and Related Photochemical Oxidants, First External Review Draft," May 5, 2011. Available as Attachment B at: http://www.regulations.gov/#!documentDetail;D=EPA-HQ-ORD-2011-0050-0009; J. E. Goodman, Comments on the 'Integrated Science Assessment of Ozone and Related Photochemical Oxidants,' EPA Document EPA/600/R-10/076A; released March 2011." Available as Attachment 1 to Docket ID EPA-HQ-ORD-2011-0050-0007.

are considered, along with the lack of biological plausibility for such serious effects, the epidemiological risk assessment should not be considered in setting the primary standard. Therefore, retention of the current standard should be considered as a viable alternative in the current review.

The following sections lay out the detailed rationale for this interpretation of the data. To provide the Administrator with a full exploration of the health effects data and range of options, the final HREA and PA should include a full discussion of these issues.

I. Comments on Background Ozone

There are two fundamental issues involving background ozone. The first is what EPA uses for background ozone which depends on how it is defined and how it is estimated. The second is how they use background in their policy and risk assessments which is then used to inform policy decisions that need to be made concerning the form, averaging time and level of the National Ambient Air Quality Standard (NAAQS). Both of these issues will be examined below.

A. The Evolution of EPA's Definition of Background Ozone

Since the release of the last Staff Paper (SP)[7] in 2007, EPA's treatment of how they consider background ozone and the role it plays in their risk and policy assessments has undergone a continuous evolution which is reflected in the changes that have occurred in the first three drafts of their Integrated Science Assessments (ISA),[8,9,10] the subsequent Health Risk and Exposure Assessments (HREA),[11,12] and the Policy Assessments (PA).[13,14] In the 2007 SP and in the first ISA draft, EPA used policy relevant background (PRB) as their preferred measure for background ozone. They defined PRB:

> The background concentrations of O_3 that are useful for risk and policy assessments informing decisions about the NAAQS are referred to as policy-relevant background (PRB) concentrations. PRB concentrations have historically been defined by EPA as

[7] U.S. EPA. 2007. *Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information OAQPS Staff Paper*, EPA-452/R-07-003.
[8] U. S. Environmental Protection Agency, *First External Review Draft of the Integrated Science Assessment for Ozone and Related Photochemical Oxidants*, EPA/600/R-10/076a, March 2011.
[9] U. S. Environmental Protection Agency, *Second External Review Draft of the Integrated Science Assessment for Ozone and Related Photochemical Oxidants*, EPA/600/R-10/076b, September 2011.
[10] U. S. Environmental Protection Agency, *Third External Review Draft of the Integrated Science Assessment for Ozone and Related Photochemical Oxidants*, EPA/600/R-10/076c, June 2012.
[11] U. S. Environmental Protection Agency, *Health Risk and Exposure Assessment for Ozone, First External Review Draft*, EPA-452/P-12-001, July 2012.
[12] U. S. Environmental Protection Agency, *Health Risk and Exposure Assessment for Ozone, Second External Review Draft*, EPA-452/P-14-004a, February 2014.
[13] U. S. Environmental Protection Agency, *Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards, First External Review Draft*, EPA–452/P–12–002, August 2012.
[14] U. S. Environmental Protection Agency, *Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards, Second External Review Draft*, EPA–452/P–14–002, January 2014.

> those concentrations that would occur in the U.S. in the absence of anthropogenic emissions in continental North America (CNA) defined here as the U.S., Canada, and Mexico. For this document, PRB concentrations include contributions from natural sources everywhere in the world and from anthropogenic sources outside CNA.[15]

The exclusion of emissions from Canada and Mexico was based on EPA's assumption that the U.S. could control emissions from Canada and Mexico by treaties and international agreements.

In the second draft of the ISA,[16] EPA stopped using the term PRB and switched to calling it North American background (NAB). EPA states: "For this document, we have focused on the sum of those background concentrations from natural sources everywhere in the world and from anthropogenic sources outside the U.S., Canada and Mexico, i.e., North American background." While they changed the term from PBR to NAB, they both had the same definition and NAB was still based on the controversial assumption that Canadian and Mexican emissions could be controlled by treaties or international agreements.

In AIR's comments[17] on the second draft of the ISA, we pointed out that their definition of NAB actually implied that Mexican and Canadian emissions could be eliminated by treaties or agreements and that this was not realistic. The way EPA used NAB resulted in their overestimating the risk reduction that would be achieved by lowering the NAAQS and it penalized the States because they would have to offset the Canadian and Mexican emissions in their State Implementation Plans. Instead of using NAB, AIR recommended that it was more appropriate to use a U.S. background (USB), which includes Canadian and Mexican emissions, for the risk assessments and for control strategy development.

In the third draft of the ISA[18], EPA finally included USB in their discussions of background. This draft included three definitions of background ozone for consideration: NAB (as previously defined), USB and natural background (NB). They define USB as the background that would exist in the absence of anthropogenic emissions from the U.S. Thus, ozone resulting from Canadian and Mexican emissions is included. EPA defines natural background as ozone "resulting from emissions from natural sources (e.g., stratospheric intrusion, wildfires, biogenic methane and more short-lived VOC emissions) throughout the globe."

In addition, the third draft also admitted the shortcomings of the "zero out" methodology EPA used to estimate NAB, USB and NB. Three times in the third draft EPA stated:

---

[15] U. S. Environmental Protection Agency, supra note 2, at pp. 2-5.
[16] U. S. Environmental Protection Agency, supra note 3, at pp. 1-4.
[17] J. M. Heuss, G. T. Wolff, and D. F. Kahlbaum, Review and Critique of the U. S. Environmental Protection Agency's Second External Review Draft of the "Integrated Science Assessment for Ozone and Related Photochemical Oxidants," Air Improvement Resource, Inc. Report, Prepared for The Alliance of Automobile Manufacturers, November 2011.
[18] U. S. Environmental Protection Agency, supra note 4, at pp. 2-7.

> Note that the calculations of background concentrations presented in this chapter were formulated to answer the question, "what would O_3 concentrations be if there were no anthropogenic sources." This is different from asking, "how much of the O_3 measured or simulated in a given area is due to background contributions." Because of potentially strong non-linearities—particularly in many urban areas—these estimates should not be used by themselves to answer the second question posed above. The extent of these non-linearities will generally depend on location and time, the strength of concentrated sources, and the nature of the chemical regime. Further work is needed on how these estimates of background concentrations can be used to help determine the contributions of background sources of O_3 to urban concentrations.

In previous comments by AIR[19] on the first draft of the ISA, this non-linearity issue was brought to EPA's attention. AIR recommended:

> The contribution of natural sources and other PRB sources to North American cannot be realistically assessed in the absence of U.S. anthropogenic emissions. To realistically estimate the contribution of PRB sources, the PRB sources should be shut down in the presence of U.S. sources.

In subsequent comments on the third draft, AIR recommended two approaches that could be used to obtain the impact of USB on US ozone concentrations:

> We suggest two approaches that should be used. In the first, the USB should be set to zero. In other words, all non-U.S. anthropogenic emissions should be set to zero as well as all of the natural sources. The difference between that scenario and the base case scenario, where all sources and emissions are included, would provide an estimate of the contribution of USB to the base case ozone. In the second approach, a photochemical grid model with an embedded source apportionment module should be used. CAMx is one such modeling system with a source-apportionment module. With USB designated as a separate source category, the estimated contribution of USB to the base case ozone would be computed directly. A comparison of the USB contributions estimated from these two approaches with the estimates of USB derived from the

[19] Wolff, G.T. 2011. Comments on Policy Relevant Background Ozone As Discussed in EPA's Draft Integrated Science Assessment for Ozone and Related Photochemical Oxidants. Prepared for the Utility Air Regulatory Group, May 5, 2011.

> approach described in the ISA would provide information of the linearity or degree on non-linearity of the system.[20]

To their credit, EPA uses CAMx to estimate USB in the second draft of the PA. Unfortunately, however, the accounting procedure that they used to track the various source contributions to USB is not appropriate for addressing the question, "how much of the O_3 measured or simulated in a given area is due to background contributions." As a result, all of the resulting USB estimates in the PA are biased low. This will be discussed in detail in the next section.

B. Background O_3 Estimates in the PA

The EPA makes extensive calculations of USB and they are contained in Chapter 2 of the PA and in Appendix 2A of the PA. They estimate USB using two different procedures: 1) by using the GEOS-Chem/CMAQ modeling system and zeroing out anthropogenic emissions in the US, and 2) by using CAMx with the Anthropogenic Precursor Culpability Assessment (APCA) tool, which has the capability of tracking the contributions of different sources (i.e.: boundary conditions, biogenic VOCs, biogenic NO_x etc.) to USB. In general, the CMAQ results in slightly higher estimates of USB than CAMx which EPA attributes to the inability of the zeroing out procedure to capture the non-linearities in the ozone chemistry. EPA states:

> While the zero-out approach has traditionally been used to estimate background ozone levels, the methodology has some acknowledged limitations. First, from a policy perspective, the purely hypothetical and ultimately unrealizable zero manmade emissions scenarios have limited application in this regard. Secondly, the assumption that background ozone is what is left after specific emissions have been removed within the model simulation can be misleading in locations where ozone chemistry is highly non-linear. Depending upon the local composition of ozone precursors, NO_x emissions reductions can either increase or decrease ozone levels in the immediate vicinity of those reductions. For those specific urban areas in which NO_x titration of ozone can be significant, zero-out modeling can result in inflated estimates of background ozone when these NO_x emissions are completely and unrealistically removed. Paradoxically, in certain times and locations in a zero-out scenario there can be more background ozone than actual ozone within the model (EPA, 2014).[21]

---

[20] J. M. Heuss, G. T. Wolff, and D. F. Kahlbaum, Review and Critique of the U. S. Environmental Protection Agency's Third External Review Draft of the "Integrated Science Assessment for Ozone and Related Photochemical Oxidants," Air Improvement Resource, Inc. Report, Prepared for The Alliance of Automobile Manufacturers, August 2012.

[21] U.S. Environmental Protection Agency, supra note 8, at p. 2A-7.

As to the application of CAMx with the APCA tool, EPA states:

> A separate modeling technique attempts to circumvent these limitations by apportioning the total ozone within the model to its contributing source terms. This basic approach is referred to as "source apportionment" modeling.

and,

> The source apportionment modeling attempts to determine how much of the modeled ozone has resulted from background sources.[22]

1. USB Estimates from CAMx

Unfortunately, the APCA tool was not designed to "determine how much of the modeled ozone has resulted from background sources," but, rather, as its name implies, to attribute maximum culpability to controllable anthropogenic precursors. As a result, the USB estimates generated by the APCA tool are negatively biased. The best way to illustrate this is with examples.

Example 1: A biogenic VOC molecule interacts with an anthropogenic NO_x molecule to form an O_3 molecule. CAMx/APCA counts the resulting O_3 as an anthropogenic O_3 molecule and not USB even though it would not exist without the contribution of the biogenic VOC. Similarly a naturally emitted NO_x molecule that reacts with an anthropogenic VOC to form O_3 would also be counted as an anthropogenic O_3 molecule.

Example 2: An O_3 molecule entering the US from Canada is initially counted as boundary condition O_3 which is included in USB. However, should that molecule encounter an NO molecule emitted from an anthropogenic source and react to form NO_2, which is subsequently photolyzed and results in the formation of an ozone molecule, that ozone molecule is counted by APCA as anthropogenic even though it would not have existed if not for the parent O_3 molecule that traveled from Canada.

Example 3: An O_3 molecule enters the US and encounters an anthropogenic VOC (olefin) which react to form an aldehyde which is subsequently photolyzed to form an ·OH radical. The ·OH radical is capable of participating in a chain reaction with NO_x and VOCs to produce many O_3 molecules. APCA would count all of the O_3 formed in this manner as anthropogenic even though they would not have existed if not for the imported initial O_3 molecule.

There are numerous other examples that could be used. For example, an O_3 molecule formed naturally in the stratosphere and transported to the troposphere would be subjected to the same accounting procedure as the ones that travel across the border into the US.

[22] Ibid.

APCA is designed to identify the controllable anthropogenic emissions that if reduced would result in lower ozone. In the case of USB, it identifies the additional reductions that would be needed to offset the USB impacts and not the contribution of USB to observed O_3. In reality, the impact of imported O_3 or NB O_3 can be propagated throughout the US by subsequent generations of reaction products that would not exist if not for the initial O_3 molecules that were transported into the US or formed naturally. In the APCA accounting procedure, the impact of these initial molecules is terminated as soon as they react with a molecule of anthropogenic origin. Thus, the CAMx/APCA modeling system underestimates the impact of USB on ambient O_3 concentrations.

2. USB Estimates from GEOS-Chem/CMAQ and Zeroing Out Anthropogenic US Emissions

As pointed out in the preceding section and admitted by EPA, the zeroing out of U.S anthropogenic emissions has some shortcomings. It alters the composition of the atmosphere including the all-important VOC to NO_x ratio that changes the chemistry in a non-linear manner that results in unrealistic consequences. One manifestation of this is that USB estimates in urban areas can be higher than the observed O_3 concentrations in the presence of U.S. anthropogenic emissions. Local NO emissions scavenge USB transported from upwind areas and this phenomena does not occur if anthropogenic NO_x emissions are turned off. In other words, the behavior and hence, the impact of USB on measured O_3 cannot be determined using this approach.

Since this procedure produces higher estimates of USB than the CAMx/APCA method, could they be considered reasonable upper estimates for USB? That is not the case because the non-linearities in the chemistry could produce higher estimates of USB when the impacts of USB are propagated throughout the US.

3. Recommended Methods to Measure Impacts of USB on Ambient Ozone Levels

As we recommended in our earlier comments, there are two ways to obtain more realistic estimates of the impacts of USB on ambient ozone concentration in the US. The first way is to run CAMx (we prefer CAMx over CMAQ because of the better performance of CAMx as was discussed in our earlier comments)[23,24] with all emissions and then with the boundary conditions and natural sources zeroed out. The difference between the two scenarios provides the impact of USB on US O_3 concentrations. The second way is to run CAMx with the APCA scheme modified to keep track of all the sources of odd oxygen atoms and distinguish between those that originated from the reactions involving

[23] Heuss et al. supra note 20.

[24] G. T. Wolff, J. M. Heuss and D. F. Kahlbaum, Review and Critique of the U. S. Environmental Protection Agency's First External Review Drafts of the "Health Risk and Exposure Assessment for Ozone" and the "Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards," Air Improvement Resource, Inc. Report, Prepared for The Alliance of Automobile Manufacturers, October 2012.

natural emissions or imported O_3 or precursors from outside the US and those formed from U.S. anthropogenic emissions alone.

4. Other Comments on Background O_3 in the PA

In Chapter 2 and Appendix 2A of the PA, EPA presents three different kinds of results: the geographical distribution of the seasonal means for USB based on the MDA8 metric,[25] the distribution of the daily estimates of MDA8 USB estimates, and the source apportionment results for the MDA8 USB estimates.

a. Seasonal Means of USB

In the PA, EPA states:

> As a first-order understanding, it is valuable to be able to characterize seasonal mean levels of background ozone. However, it is well established that background levels can vary substantially from day-to-day. From an implementation perspective, the values of background ozone on possible exceedance days is a more meaningful distinction.

We agree completely with this statement. This statement represents a significant departure in the Agency's approach to the treatment of background O_3 from the previous review where seasonal or monthly means were used in control strategy discussions and to calculate risk estimates.[26] We feel the presentation of the geographical distribution of the seasonal means of MDA8 USB is useful in gaining an understanding of the climatology of O_3. However, it is not relevant to the NAAQS which is based on an extreme value 98-percentile statistic. In addition, it should be kept in mind that the MDA8 USB estimates from CAMx that are used in the discussion of Figures 2-12 and 4d are biased low for the reasons discussed in section I.B.1 above.

b. Distributions of Daily USB

The discussion of the distributions of the daily MDA8 USB concentrations are more relevant to the discussion of the O_3 NAAQS. The focus of EPA's discussion in the PA is on Figures 2-13 and 2-14, which have been reproduced below as Figures 1 and 2. Figure 1 shows the absolute estimates of MDA8 USB as a function of total modeled MDA8 while Figure 2 plots the same data divided by total MDA8. The explanation of the boxes and whiskers are:

> a. the median concentration (black horizontal line) per bin,
> b. the inter-quartile range (blue colored box) which represents the 25th-75th percentile range in values within the distribution,

[25] MDA8 is the maximum daily 8-hour ozone concentration.
[26] Environmental Protection Agency, supra note 1.

> c. the "whiskers" (dark gray vertical lines with top and bottom whiskers) which represent the range of values within 1.5 times the inter-quartile range, and
> d. the "outliers" (gray points) which are any values outside the whiskers.[27]

EPA's main conclusion from their analysis is that it supports their earlier conclusions that:

> "results suggest that background concentrations on the days with the highest total O_3 concentrations are not dramatically higher than typical seasonal average background concentrations." Based on this finding, EPA determined that "anthropogenic sources within the U.S. are largely responsible for 4th highest 8- hour daily maximum O_3 concentrations."[28]

and,

> For example, for site-days in which base O_3 is between 70-75 ppb, the source apportionment modeling estimates that approximately 37 percent of those O_3 levels originate from sources other than U.S. anthropogenic emissions (i.e., apportionment-based USB).[29]

These statements cannot be disputed if the caveat "on average" is inserted. But the O_3 NAAQS is not about "on average" occurrences. The O_3 NAAQS is about events that cause extreme values. Since the NAAQS is based on the annual 98th-percentile, EPA should have identified the 98th-percentile value on the whisker plots. In addition, they should provide plots for each individual city used in the risk assessments. Clearly Figures 1 and 2 show that there are times when the O_3 is within the range of the values being considered for the NAAQS that USB contributes significantly to the modeled MDA8 O_3 concentration and in a few cases is sufficient to cause an exceedance by itself. Concerning these rare events, EPA states:

> Figure 2-14 also indicates that there are cases in which the model predicts much larger background proportions, as shown by the upper outliers in the figure. These infrequent episodes usually occur in relation to a specific event, and occur more often in specific geographical locations, such as at high elevations or wildfire prone areas during the local dry season.

[27] Environmental Protection Agency, supra note 8, at p. 2A-21.
[28] Ibid, p. 2-16.
[29] Ibid, p. 2-17.

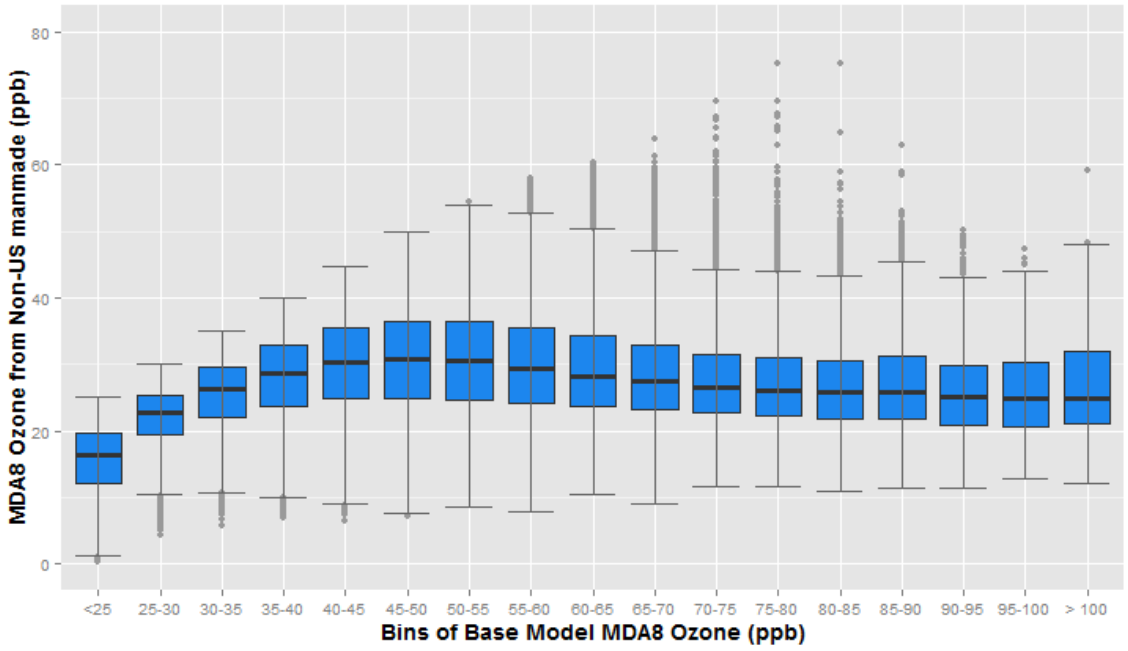

Figure 1: Distribution of MDA8 ozone contributions from non-U.S. manmade sources (USB) in ppb at monitoring locations across the U.S. (Apr-Oct), binned by base modeled site-day MDA8, as estimated by 2007 CAMx simulations.

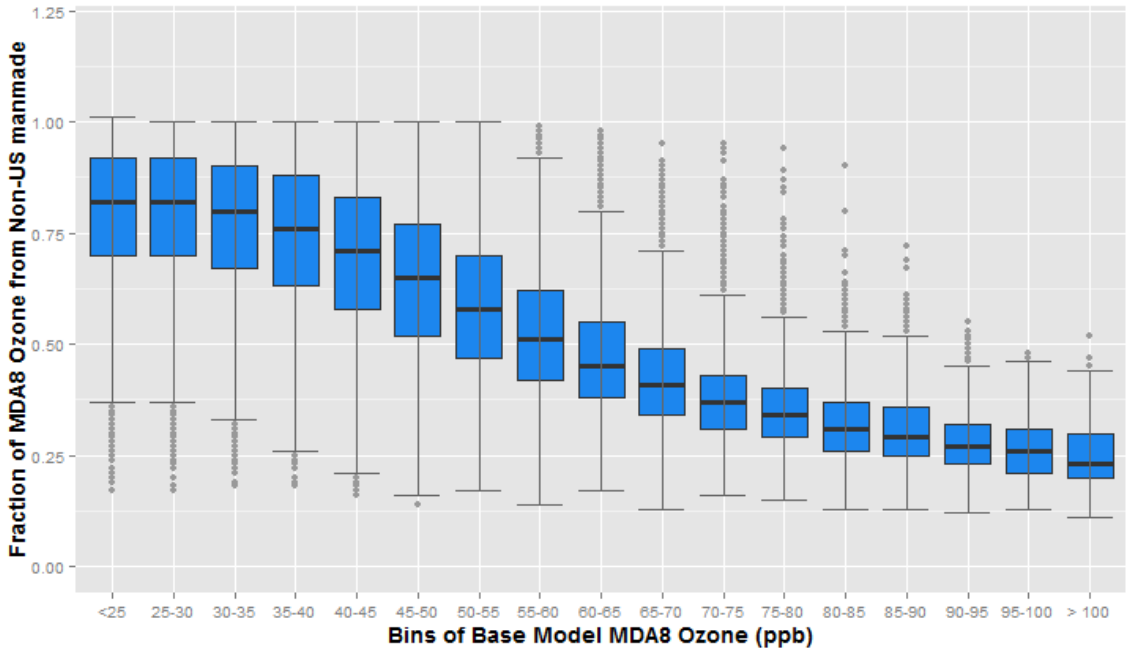

Figure 2: Distribution of MDA8 ozone fractions from non-U.S. anthropogenic sources (USB) at monitoring locations across the U.S. (Apr-Oct), binned by base modeled site-day MDA8, as estimated by the 2007 CAMx simulation.

> It should be noted here that EPA has policies for treatment of air quality monitoring data affected by these types of events. EPA's exceptional events policy allows exclusion of certain air quality monitoring data from regulatory determinations if a State adequately demonstrates that an exceptional event has caused the exceedance or violation of a NAAQS. In addition, Section 179B of the CAA also provides for treatment of air quality data from international transport when an exceedance or violation of a NAAQS would not have occurred but for emissions emanating from outside of the United States.[30]

EPA gives the impression that an exceedance of the NAAQS being caused by a contribution from background ozone is a rare event and should it occur, they have mechanisms in place that eliminate consideration of the event in the determination of an areas attainment/nonattainment status. Since 2008, EPA has only considered and approved two events for O_3[31] as qualifying for "exceptional event status," and they made the states jump over some very high hurdles to make their case. The first one was associated with the impact of numerous wildfires on an O_3 monitor in the Sacramento area in the summer of 2008. It took the California Air Resources Board (CARB) almost two and a half years with the help of consultants to assemble massive documentation and analyses in a 85 page report plus nine appendices that was submitted to EPA in March, 2011.[32] In the second case, O_3 monitors in the Wichita, KA were being impacted by wildfires in Kansas, Oklahoma and Texas during April, 2012. With the help of contractors, the Kansas Department of Health and Environment submitted a 247 page report to EPA in November, 2012.[33] Obviously, these states had to expend considerable monetary and personnel resources to produce the required documentation.

Thus far, these two wildfire incidents have been the only two exceptional event cases for O_3. There have been no cases involving international transport even though that is happening to some extent every day. On their Exceptional Event website,[34] EPA has a link to a presentation by Neil Frank.[35] One slide contains the following information:

> Rule Requirements
> – Event satisfies the definition of exceptional
> – There is a clear causal relationship

---

[30] Ibid.
[31] U.S. EPA. 2014. Treatment of Data Influenced by Exceptional Events. http://www.epa.gov/ttn/analysis/exevents.htm. (Accessed February 24, 2014).
[32] CARB. 2011. Exceptional Events Demonstration for 1-Hour Ozone Exceedances in the Sacramento Regional Nonattainment Area Due to 2008 Wildfires. http://www.epa.gov/ttn/analysis/exevents.htm. (Accessed February 24, 2014).
[33] Kansas Department of Health and Environment Division of Environment Bureau of Air. 2012. Exceptional Event Demonstration Package April 6, 12, 13, and 29, 2011. http://www.epa.gov/ttn/analysis/docs/KDHE_ExEvents_final_042011.pdf (Accessed February 24, 2014).
[34] U.S. EPA. 2014. supra note 31.
[35] N. Frank. 2009. Presenting Evidence to Justify Data Exclusion as an Exceptional Event Exclusion as an Exceptional Event. http://www.epa.gov/ttn/analysis/docs/IdeasforShowingEEEvidence.pdf. (Accessed February 24, 2014).

– Event is associated with measured concentration in excess of normal fluctuations including background
– No exceedance or violation but for the event

The third statement on this slide is sufficiently ambiguous that EPA could disqualify most exceedances where USB provides a major contribution. It would be helpful if EPA would define "normal fluctuations including background" and illustrate it in Figure 1.

c. Source Apportionment Derived USB in 12 Urban Areas

In this section of the PA, EPA uses CAMx and the APCA tool to derive various mean estimates of USB for the 12 urban areas that EPA uses as case study areas in the HREA. Even though these USB estimates are biased low for the reasons discussed above, it is useful to note the large fraction of the modeled MDA8 O_3 that is due to USB. Average estimates of USB on all days are presented in the PA in Tables 2-1 and 2-3 (reproduced as Tables 1 and 2 below) for all 12 of the cities. The average fraction of the MDA8 O_3 due to USB range from 0.43 in Atlanta to 0.69 in Denver. The average of all the cities is about 0.6. This is a non-trivial contribution. On average, most of the MDA8 O_3 in most of the cities is from USB. How much does this fraction have to vary from the mean to be considered an exceptional event?

EPA also estimates the fraction due to USB on days when the MDA8 O_3 60 ppb or greater in Table 3. On these days, the fraction due to USB range from 0.31 in Baltimore and Philadelphia to 0.55 in Denver. These fractions are hardly non-trivial. It would have been useful if EPA generated plots of the daily distributions (like in Figure 1 and 2) for each of the 12 cities.

All days, CAMx	ATL	BAL	BOS	CLE	DEN	DET	HOU	LA	NYC	PHI	SAC	STL
Model MDA8 seasonal mean	59.3	54.4	43.0	48.9	47.3	39.1	48.5	51.1	45.4	48.7	46.4	49.8
Model MDA8 seasonal mean from emissions other than U.S. anthropogenic sources	25.3	25.9	26.2	25.7	31.3	23.3	27.0	29.1	24.5	24.2	29.7	24.3
Fractional contribution from background	0.43	0.48	0.61	0.52	0.66	0.60	0.56	0.57	0.54	0.50	0.64	0.49

Table 1. Seasonal mean MDA8 O_3 (ppb), seasonal mean apportionment-based USB contribution (ppb), and fractional apportionment-based USB (CAMx) contribution to total O3 (all site-days) in the 12 REA urban case study areas (%).

	ATL	BAL	BOS	CLE	DEN	DET	HOU	LA	NYC	PHI	SAC	STL
Mean of daily MDA8 background fractions	0.46	0.53	0.68	0.58	0.69	0.64	0.59	0.61	0.61	0.56	0.67	0.52
Median of daily MDA8 background fractions	0.43	0.51	0.73	0.54	0.69	0.66	0.59	0.60	0.63	0.54	0.66	0.49

Table 2. Fractional contribution of apportionment-based USB in the 12 REA urban study areas (%), using the means and medians of daily MDA8 fractions (instead of fractions of seasonal means).

Only days w/ base MDA8 > 60 ppb	ATL	BAL	BOS	CLE	DEN	DET	HOU	LA	NYC	PHI	SAC	STL
Model MDA8 seasonal mean	74.0	75.3	70.7	72.0	67.5	68.9	70.3	74.4	74.1	74.0	68.3	70.0
Model MDA8 seasonal mean from emissions other than U.S. anthropogenic sources	25.4	23.7	24.4	25.4	37.3	24.4	28.0	31.9	23.5	22.9	32.1	25.4
Fractional contribution from background	0.34	0.31	0.35	0.35	0.55	0.35	0.40	0.43	0.32	0.31	0.47	0.36

Table 3. Seasonal mean MDA8 O_3 (ppb), seasonal mean apportionment-based USB contribution (ppb), and fractional apportionment-based USB contribution to total O_3 (site-days > 60 ppb) in the 12 REA urban study areas (%).

C. Use of Background O_3 in the HREA

In the previous O_3 NAAQS review, EPA estimated O_3 exposure risks down to background O_3, which, at the time, they defined as PRB.[36] EPA was criticized for the use of PRB because it did not contain Canadian and Mexican contributions and EPA used mean monthly values. To avoid such criticisms in the present review, EPA has decided to calculate risks that are independent of the choice of background. In the second draft HREA, risks are estimated down to a concentration of zero ppb. By doing this, EPA has inflated the risk estimates. For the reasons discussed in the health effects section of these comments, EPA health risks are not realistic.

Because USB contributes significantly to MDA8 O_3 and because EPA does not exclude USB levels in their risk assessments, additional anthropogenic emission reductions will be needed to offset the impact of USB. As a result, extreme additional emission reductions will be required to achieve the alternative NAAQS being considered by EPA.

In previous risk assessments, EPA employed the quadratic rollback method to estimate the spatial and frequency distributions of MDA8 O_3 in urban areas after just meeting any alternative NAAQS. As EPA has articulated, there are a number limitations with this approach. As a result, EPA uses a new method in the second draft REA that we believe is a superior method. That method is to use the Community Multi-scale Air Quality photochemical model (CMAQ)[37] in conjunction with the Higher order Direct Decouple Method (HDDM) to estimate the distributions of MDA8 O_3 concentrations associated with achieving alternative NAAQS. At the same time, the method estimates the degree of emission controls that the urban areas need to apply in order to achieve alternative NAAQS. EPA has tabulated the required emission reductions for 15 cities for two base periods, 2006-2008 and 2008-2010 and these have been reproduced in Table 4.[38] Although EPA qualifies these estimated emission reductions,

[36] U.S. EPA, supra note 7.
[37] Because of superior model performance of CAMx relative to CMAQ, we would recommend that CAMx be used instead of CMAQ.
[38] U.S. EPA. 2014. Health Risk and Exposure Assessment for Ozone, Second External Review Draft, Chapter 4 Appendices. EPA-452/P-14-004b. pp. 26-27.

> Please note that these reductions and broad nationwide emission cuts are not intended to represent recommended control scenarios since they would not be the most efficient method for achieving the standard in many localized areas.

they are certainly useful in providing "ball park" estimates of the emission reductions required to achieve the alternative NAAQS. To achieve a 70 ppb NAAQS, NO_x reductions of 27 - 89% are needed and to achieve a 60 ppb NAAQS, NO_x reductions of 62 - 93% are needed.

Using CAMx and HDDM and a base year of 2006, Downey et al., 2014[39] made similar reduction estimates for both VOC and NO_x emissions. Their estimates are shown in Table 5. These estimates are slightly greater than EPA's, but nevertheless are similar in magnitude.

Both EPA's and Downey et al.'s estimates are consistent and indicate that massive additional anthropogenic emission reductions are going to be required to meet any of the alternative NAAQS levels that are being considered by the EPA.

D. Summary of Background O_3 Discussion

In the Second Draft PA, EPA finally is using USB as the relevant measure of U.S. background O_3 in their discussion on policy related issues. However, the two methods they use to estimate USB, zeroing out U.S. anthropogenic emissions in the CMAQ modeling system and running CAMx with the APCA tool, do not provide realistic estimates of the impact of USB on modeled daily MDA8 O_3 concentrations. We have provided the reasons for this and have suggested two alternative approaches that will provide more realistic estimates of the impacts of USB on MDA8 O_3.

Even though EPA underestimates USB, their analyses still illustrate an important contribution of USB to the modeled MDA8 concentrations. On average, USB is responsible for the majority of

the observed and modeled MDA8 O_3 concentrations throughout the U.S. On high O_3 days (MDA8 $\geq$ 60 ppb), the USB still contributes significantly and, on occasion, can provide all or most of the O_3.

EPA has in place a procedure to identify and eliminate from attainment/nonattainment considerations "exceptional events" caused by high USB. However, this procedure has only been used twice since 2008 for O_3 from wildfires and requires a considerable time and resource investment by the states involved. Further, it does nothing to ease the burden on the states from the daily persistent contribution from USB.

---

[39] Downey, N.; C. Emery; J. Jung; T. Sakulyanontvittaya; L. Hebert; D. Blewitt; G. Yarwood. 2014. "Emissions Reductions and Urban Ozone Responses under More Stringent US Standards. Environ. Sci. Technol. (Submitted).

EPA's decision to estimate risks to O_3 exposures down to zero ppb instead of to USB increases the calculated risks due to exposures that are not a result of U.S. anthropogenic emissions. To offset the impact of USB, extreme additional reductions of U.S. anthropogenic emissions will be required to achieve the alternative NAAQS being considered by EPA.

Urban Area	Years	Standard Level*			
		75 ppb	70 ppb	65 ppb	60 ppb
Atlanta	2006-2008	50%	58%	64%	71%
	2008-2010	23%	43%	54%	62%
Baltimore	2006-2008	46%	54%	61%	69%
	2008-2010	44%	52%	60%	67%
Boston	2006-2008	40%	49%	61%	70%
	2008-2010	13%	40%	53%	65%
Chicago	2006-2008	19%	52%	66%	80%
	2008-2010	N/A	27%	55%	70%
Cleveland	2006-2008	48%	61%	73%	88%
	2008-2010	50%	64%	77%	88%
Dallas	2006-2008	50%	57%	65%	72%
	2008-2010	50%	58%	64%	71%
Denver	2006-2008	51%	65%	78%	87%
	2008-2010	15%	46%	64%	87%
Detroit	2006-2008	59%	69%	76%	84%
	2008-2010	N/A	54%	66%	78%
Houston	2006-2008	62%	68%	74%	82%
	2008-2010	42%	53%	63%	75%
Los Angeles	2006-2008	87.1%	89.3%	91.2%	93.2%
	2008-2010	87%	89%	91%	93%
New York	2006-2008	64%	74%	92%	N/A
	2008-2010	52%	67%	89%	N/A
Philadelphia	2006-2008	54%	61%	68%	74%
	2008-2010	42%	52%	61%	68%
Sacramento	2006-2008	63%	70%	76%	84%[$]
	2008-2010	64%	71%	77%	84%
Saint Louis	2006-2008	45%	56%	66%	75%
	2008-2010	10%	34%	50%	63%
Washington D.C.	2006-2008	53%	60%	67%	74%
	2008-2010	31%	50%	60%	71%

* N/A values for the 75 ppb standard level mean that a particular urban area did not have any design values above 75 for that 3-year period so no controls were needed. N/A values for the 60 ppb standard level mean that this adjustment methodology was not able to bring design values down to 60 for that particular city and 3-year period.

Table 4. Percent emission reductions used for each urban area to achieve each alternative NAAQS. Percentages in Chicago and Denver represent reductions in both anthropogenic NO_x and VOC. Percentages in all other cities represent reductions in NO_x only.

	% Reductions in 2006 Emissions			
4th highest MDA8	Los Angeles, CA	Sacramento, CA	St. Louis, MO	Philadelphia, PA
75 ppb	92%	68%	62%	62%
70 ppb	94%	75%	70%	73%
65 ppb	95%	82%	77%	79%
60 ppb	97%	87%	85%	84%

Table 5. Percent emission reductions for both VOC and NO_x needed to achieve alternative NAAQS (from Downey et al., 2014).

II. Comments on the HREA

The stated goal of the HREA is to provide information relevant to answering questions regarding the adequacy of the existing O_3 standard and the potential improvements in public health from meeting alternative standards.

To achieve this goal, the HREA presents analyses that provide:

> (1) estimates of the number of people in the general population and in at-risk populations and lifestages with O_3 exposures above benchmark levels, while at moderate or greater exertion levels;
> (2) estimates of the number of people in the general population and in at-risk populations and lifestages with impaired lung function resulting from exposures to O_3; and
> (3) estimates of the potential magnitude of premature mortality and selected morbidity health effects in the population, including at-risk populations and lifestages, where data are available to assess these groups.[40]

In addition, the HREA includes an effort to meet the following additional goals:

> 4) to evaluate the influence of various inputs and assumptions on risk estimates to the extent possible given available methods and data;
> (5) to gain insights into the spatial and temporal distribution of risks and patterns of risk reduction and uncertainties in those risk estimates.

The HREA builds upon and uses much the same methodology as that used in the last review completed in 2008. Some changes and improvements were made for the first draft HREA and additional changes have been made to the methodology for the second draft HREA. AIR has reviewed the second draft with the goals in mind and has comments on each of the major analyses and goals.

A. Ozone exposures above benchmark levels with exercise

The first listed goal is to provide estimates of the number of people with O_3 exposures with moderate or greater exercise above benchmark levels. This calculation, however, is

[40] HREA, supra note 1, at p. 3-5.

not directly a measure of risk of adverse effects or risk to public health. Although the benchmarks chosen -- 8-hour exposures of >60 ppb, >70 ppb and >80 ppb -- coincide with the concentrations used in the most recent clinical studies, this calculation does not include consideration of any physiological responses. In addition, the physiological responses from single exposures to such levels have not been considered adverse in prior reviews.

The role of exercise in eliciting the first physiological effects of ozone is particularly important. It should be borne in mind that a subject has to be outside, exercising at the time and place of high ozone for there to be an exposure that could cause an effect. In order to calculate such exposures, all these factors need to be taken into account and this is what the APEX (Air Pollution Exposure Model) attempts to do. Because of the importance of exercise, the portions of the model that simulate activity and ventilation rate need special scrutiny.

The presentation of the output of the headcount analysis in the HREA is misleading and not directly relevant to public health. The HREA notes that APEX provides two basic outputs (1) counts of people exposed one or more times to a given O_3 concentration while at a specified breathing rate, and (2) counts of person-occurrences which accumulate occurrences of specific exposure conditions over all people in the population groups of interest over an ozone season. The first of these metrics, counts of people exposed one or more times, is not as relevant to public health as the second metric, counts of person – occurrences over the entire group and ozone season. Single occurrences of small, transient FEV1 decrements have not been considered adverse during prior reviews, so being exposed only once a season is not particularly relevant to public health. On the other hand, the second metric can be quite informative of the portion of people and portion of time when there may be potential risk.

Despite the inclusion of much information on the distribution of person-occurrences in the APEX output and its more direct relevance to public health, the HREA focuses on the first metric. There are multiple figures and tables in Chapter 5 and extensive discussion of the results for the first metric and no presentation at all for the second metric. This must be remedied in the final HREA.

AIR made this point in comments on the first draft HREA, and presented two examples based on sample applications for APEX 4.5 on the EPA website.[41] One example was for a 2006 Denver base case. In 2006, the ozone design value for Denver was 90 ppb, which is greater than the current ozone standard of 75 ppb. The Denver simulation estimated the distribution of exposures for 550,471 children for 204 days, for a total of 1.123×10^8 total person-days. For all exposures without regard to exertion level, the APEX application predicted that only 0.004 or 0.4 percent of the children's 8-hour exposures are 60 ppb or greater. For all exposures at 13 EVR or greater, APEX predicted that only 0.0027 or 0.27 % occur at 8-hour exposures of 60 ppb or greater. For the cutpoint of 70 ppb, the portion of maximum 8-hour exposures with EVR of 13 or greater was 0.00057 or 0.057 %. Thus, in the 2006 base case, the vast majority of children's exposures are

[41] AIR comments, supra note 5, at p. x.

below the level of any concern. Attainment of the current standard would reduce the already extremely small portion of exposures substantially. To provide a more complete perspective on the impact of the current and alternative standards, the HREA must include presentation of both persons and person-days exposure results. Since the HREA focuses on the percent of subjects experiencing one or more benchmark exposures a year, the overall percent of person-days above the benchmark should also be presented for a complete picture of the APEX output.

Because of the importance of exercise, the portions of the APEX model that simulate activity and ventilation rate need special scrutiny. Because of the small portion of total person-days above the various benchmarks, the extremes of the predictions from APEX need special scrutiny, too. AIR has identified three ways in which the estimates of benchmark exposures in Chapter 5 the draft HREA are biased high. These concerns were identified to the Agency during the prior review[42] and re-iterated with regard to the first draft HREA.[43] In all three cases, the Agency has been aware of the concerns, either acknowledges the concern or presents data to confirm the concern in the HREA, yet has chosen not to evaluate the sensitivity of the results to these factors.

First, the APEX model predicts more elevated ventilation rate occurrences than observed in real world data. In the previous review, Langstaff acknowledged that the "values produced by the ventilation rate algorithm may exhibit an excessive degree of variability."[44] An excessive degree of variability will produce an excessive number of extreme values of ventilation rate.

The 1997 EPA analysis had also over-estimated the number of high ventilation rates in the population by using an algorithm to assign ventilation rates based on individuals who exercised regularly and were motivated to reach a high ventilation rate. As a result, the 1996 Staff Paper acknowledged that the analysis allowed more high ventilation rates (hence greater risk) than would actually occur in the populations of interest - outdoor workers, outdoor children, etc.[45]

The final sensitivity analysis for APEX in the previous review included a comparison of predicted ventilation rates with mean values in the literature, but the upper tails of the distribution which impact the risk estimates were not compared.[46] This was an important oversight because the upper percentiles of ventilation rate are responsible for the exposures that cause the perceived risk.

In comments on the first draft HREA, AIR presented a comparison of the APEX modeled

---

[42] Alliance comments, supra note 3, at p. 13-16.
[43] AIR comments, supra note 5, at p. 24-28.
[44] J. Langstaff Technical Memorandum, *Analysis of Uncertainty in Ozone Population Exposure Modeling*, Jan. 31, 2007 at pp.42 (EPA-HQ-OAR-2005-0172-0174).
[45] U. S. Environmental Protection Agency, *Review of the National Ambient Air Quality Standards for Ozone: Assessment of the Scientific and Technical Information*, OAQPS Staff Paper, EPA-452/R-96-007, June 1996, pp. 62-72.
[46] Ibid., at pp. 52.

values with the measured ventilation rates from Brochu et al. (2006),[47] in which the model over-predicted mean daily ventilation rates for persons below age 11 and over age 40 and, more importantly, the model had a much higher standard deviation at all ages. The second draft addresses the mean ventilation rates and makes a correction to the Brochu et al. data to show closer agreement with the APEX results, concluding that "this overall agreement suggests reasonable confidence can be conferred to the algorithm used by APEX to estimate, at a minimum, daily mean ventilation rates."[48] However, the second draft does not address the difference in variability around the mean, which is the key issue concerning the extreme values of ventilation rate. This suggests that the upper percentiles of ventilation rates in the model are substantially above those measured by Brochu et al. in a database of over 30,000 person-days from a cohort of over 2,200 free-living individuals between the ages of 3 and 96.

There is, however, a new evaluation of APEX in the HREA that can provide further insight into this issue. The analysis discussed in Appendix 5G-5 and summarized at page 5-49 of the HREA was performed using a subset of personal O_3 exposure measurements obtained from the Detroit Exposure and Aerosol Research Study (DEARS).2012). For five consecutive days, personal O_3 outdoor concentrations along with daily time-location activity diaries were collected from 36 adult study participants in Wayne County Michigan during July and August 2006. An APEX simulation was performed considering these same geographic and temporal features. Although the outdoor concentrations and time outdoors tracked well between the simulation and the observations, there were major differences in the mean daily ozone exposures and, importantly, the maximum daily ozone exposures, as shown in Figure 5-15 from the HREA, reproduced below as Figure 3. This evaluation clearly shows the influence of the excessive variability in the APEX model.

[47] P. Brochu, J. Ducre-Robitaille, and J. Brodeur, Physiological daily inhalation rates for free-living individuals aged 2.6 months to 96 years based on doubly labeled water measurements: comparison with time-activity- ventilation and metabolic energy conversion estimates, *Int. J. Hum. Ecol. Risk. Asses*., 12, 736-761 (2006).

[48] HREA, supra note 1, at 5-54.

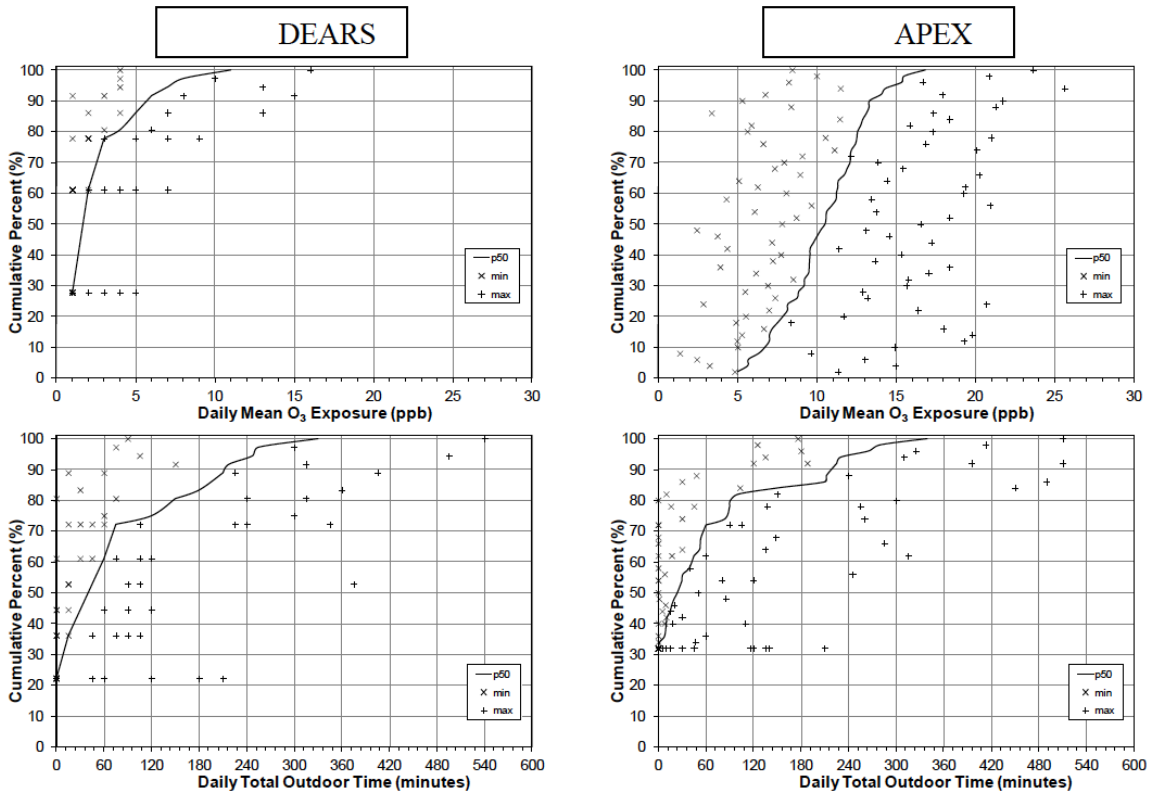

Figure 3. Distribution of daily average O_3 exposures (top panels) and daily afternoon outdoor time (bottom panels) and for DEARS study participants (left panels) and APEX simulated individuals (right panels) in Wayne County, MI, July-August 2006.

The HREA highlights the difference in the mean daily ozone exposure between the observations and the model, indicating that it is unexpected because of the distinct matching of influential personal attributes in the comparison. Although the HREA indicates the reason(s) for the difference is being investigated, the document is silent on the major difference in the extremes. The excessive variability in the ventilation rate algorithm(s) used by EPA in successive ozone reviews has been known and acknowledged for over 25 years. In addition to the prior acknowledgements documented above, the HREA indicates "APEX estimated daily ventilation rates can be greater (2-3 m3/day) than literature reported measurement values (Table 25 of Langstaff, 2007)" and that "Also, a shorter-term comparison (hours rather than daily), while more informative, cannot be performed due to lack of data."[49]

There are other datasets available to evaluate the APEX output. For example, the HREA acknowledges "the range in percent of outdoor time associated with strenuous activities using the CHAD asthmatic diaries extends beyond that of asthmatic persons from the three independent studies by about a factor of two higher."[50] In addition, Shamoo et al., 1991[51] investigated the summer activity patterns of outdoor workers in Los Angeles and

[49] HREA, supra note 1, at 5-64.
[50] HREA, supra note 1, at p. 5-40.
[51] D. Shamoo, T. Johnson, S. Trim, D. Little,W. Linn, and J. Hackney, Activity patterns in a panel of outdoor workers exposed to oxidant pollution, *J. Exp. Anal. Environ. Epidemiol*., 1, 423-438 (1991).

reported estimated ventilation rates based on heart rate recordings. The subjects also used diaries to record their location and activity. The ventilation rate reported for fast activity (44 L/min) was comparable to the ventilation rate used in the recent clinical studies. The outdoor workers diaries showed fast activity only 1 % of the time, and only at leisure, never at work.

Thus, there are several datasets that indicate a bias in the upper extremes of ventilation rate that determine the exceedances of the benchmark levels in the risk assessment. Even without more data, it would be straightforward to evaluate the sensitivity of the exposure portion of the risk assessment to the variability in the ventilation rate algorithm. The final HREA should include a specific sensitivity calculation.

A second way the counts of benchmark exposures are biased high relates to how EPA defines moderate or greater exercise over 8 hours. The HREA follows the approach begun in 1996 of defining Equivalent Ventilation Rates (EVRs) between 13 and 27 as moderate.[52] The counts in Chapter 5 thus accumulate exposures accompanied by 8-hour EVRs of 13 or greater. In Chapter 6, the risks are calculated for individuals with daily 8-hour average EVR greater than 13 using response functions developed from chamber study data conducted at a significantly higher EVR, ~ 20. In AIR comments on the first draft HREA, we presented data generated by Ted Johnson that showed the EPA algorithm predicts that the 95th percentile 8-hour EVR is between 14 and 15 while the EVR used in the clinical studies of 20 is about the 99th percentile.[53] We included figures showing the distribution of mean EVR, maximum 2-hour EVR and maximum 8-hour EVR for both asthmatics and non-asthmatics. We noted that APEX accumulates headcounts for subjects that are associated with 8-hour EVRs in the low 90s of percentiles while the EVR used in the clinical studies represents the 99th percentile. Thus, the resulting benchmark headcounts overestimate the number of subjects at potential risk in Chapter 5 and the resulting risks calculated with the EVR method in Chapter 6 are unreasonably high. The figures are reproduced below as Figures 4 and 5.

[52] HREA, supra note 1, at pp. 5-16.

[53] T. Johnson, "Background Information on EVR Sequence Statistics, September 25, 2007, Attachment 2 to Comments of the American Petroleum Institute on National Ambient Air Quality Standards for Ozone, Proposed Rule, October 9, 2007, Docket No. EPA-HQ-OAR-2005-0172-12158-1.1.

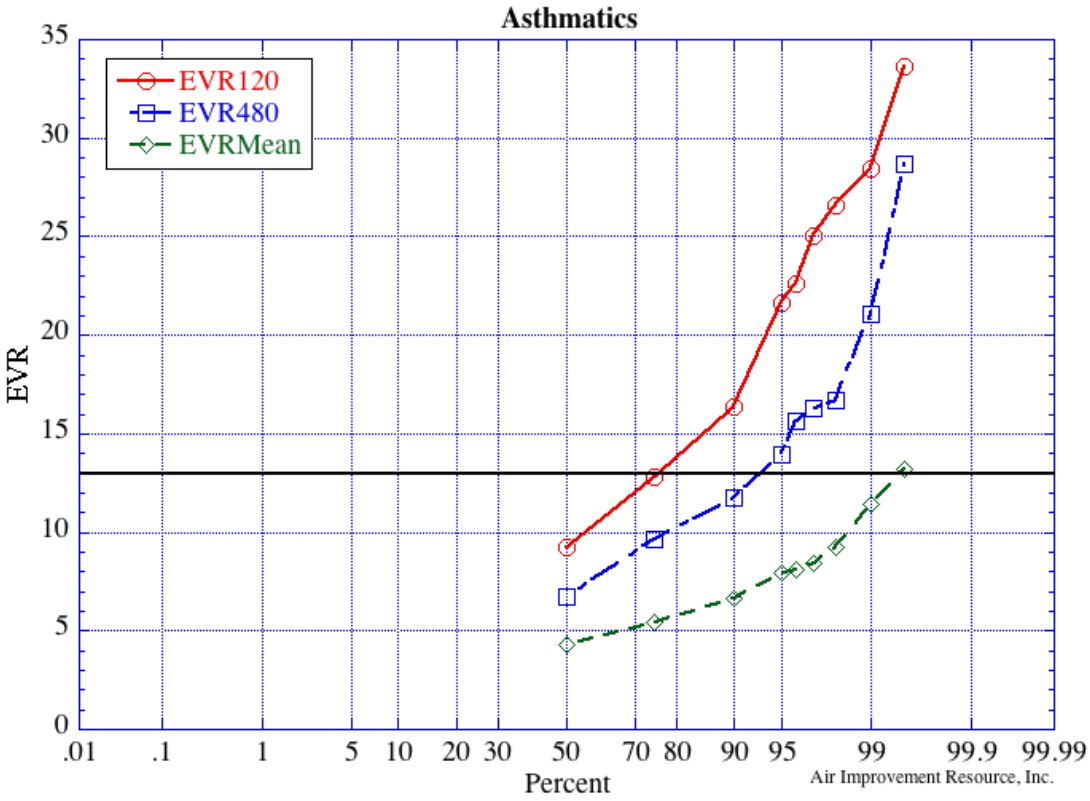

Figure 4. Distribution of EVRs calculated by the APEX algorithm for asthmatics

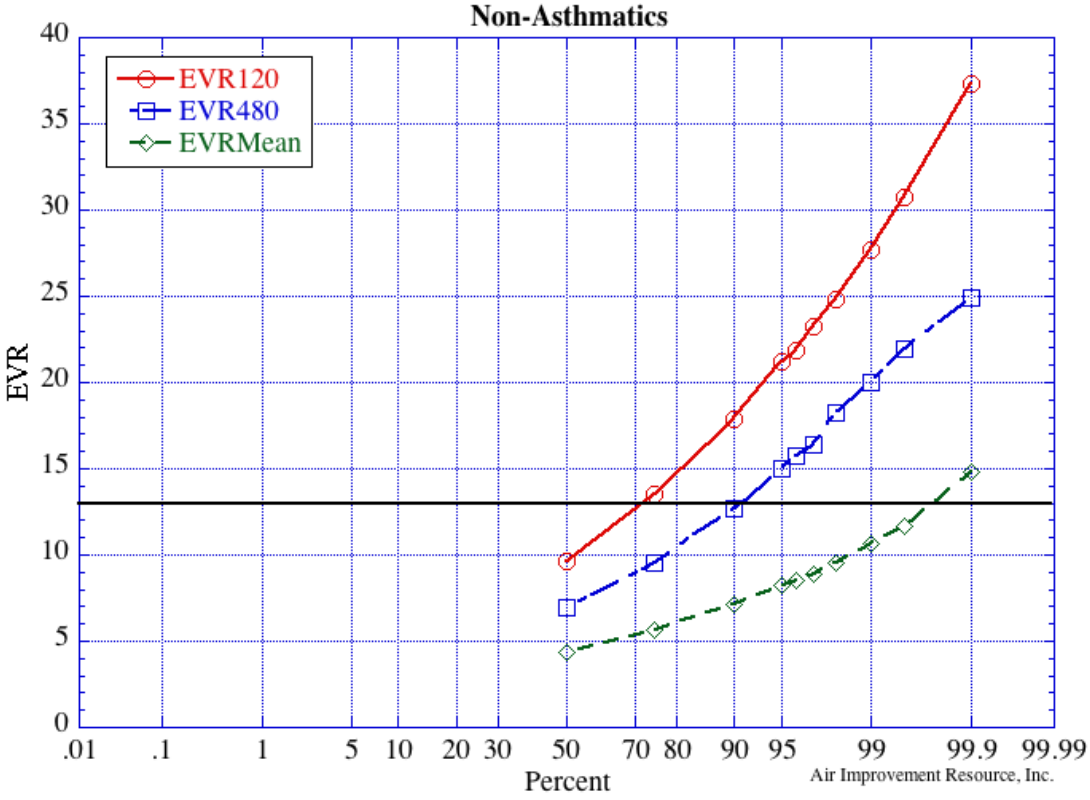

Figure 5. Distribution of EVRs calculated by the APEX algorithm for non-asthmatics

The various new studies of exposure to 0.060 ppm while exercising all utilize an experimental protocol that is quite strenuous compared to the normal range of human activity. In the Kim et al. (2011) study, the heart rate of the subjects with either ozone or filtered air averaged 127 or 128 beats per minute over the 6.6-hour test period. This means that the heart rate was higher during the six 50-minute exercise periods. While such a heart rate is common with exercise, it is not common to exercise at such a rate for such a long time. In fact, it is not unlike the heart rate achieved by a typical marathon runner who runs at between 70 and 80 % of their maximum heart rate, typically 135 beats per minute, for most of the race.

In addition, Schelegle et al. (2009) point out that the mean overall ventilation used in their study is equal to or greater than mean ventilations that might be encountered during a day of heavy to severe manual labor among the construction workers observed by Linn and colleagues[54] and that this represents the higher end of ventilations that might be encountered in the normal population for this prolonged period. Schelegle et al. recruited subjects that were engaged in a regular program of aerobic training to ensure their ability to complete the exercise protocol which was five exposure scenarios with a minimum of seven days between exposures. Nevertheless, a CASAC panelist in preliminary comments noted that less than half the subjects completed the 6.6 hour exposure protocols.[55] Thus, there is a mismatch between the strenuous protocols used in the recent clinical studies and the >13 EVR cutpoint used in the headcount analysis.

The HREA acknowledges the mis-match, noting:

> Given that the EVR serves as a cut point for selecting persons performing at moderate or greater exertion and is a lower bound value (~5th percentile), the simulated number of persons achieving this level of exercise is possibly overestimated.[56]

The HREA in Chapter 5 notes that this is a newly identified concern and that it "may need additional characterization." It is not a newly identified concern as AIR pointed this out several times in the prior review and in the current review.

In addition, Chapter 6 includes data demonstrating the concern that AIR has raised. Figure 6-11 in the HREA, reproduced below as Figure 6, shows that the distribution of EVRs greater or equal to 13 for the Atlanta simulation

> …is clearly shifted much lower than the distribution of EVR in the clinical studies. This could lead to an overestimation of the percent of responders by the E-R method, since higher EVRs lead to higher lung function decrements and it is

[54] W. Linn, C. Spicer, and J. Hackney, "Activity patterns in ozone-exposed construction workers," *J. Occup. Med. Toxicol*., 2, 1-14 (1993).
[55] Preliminary Individual Comments on Health Risk and Exposure Assessment for Ozone (First External Review Draft, Updated August 2012), from members of the CASAC Ozone Review Panel, dated September 4, 2012, at pp. 32.
[56] HREA, supra note 1, at 5-64.

applying an E-R function based on EVRs around 20 to a population with median EVRs around 14.5.[57]

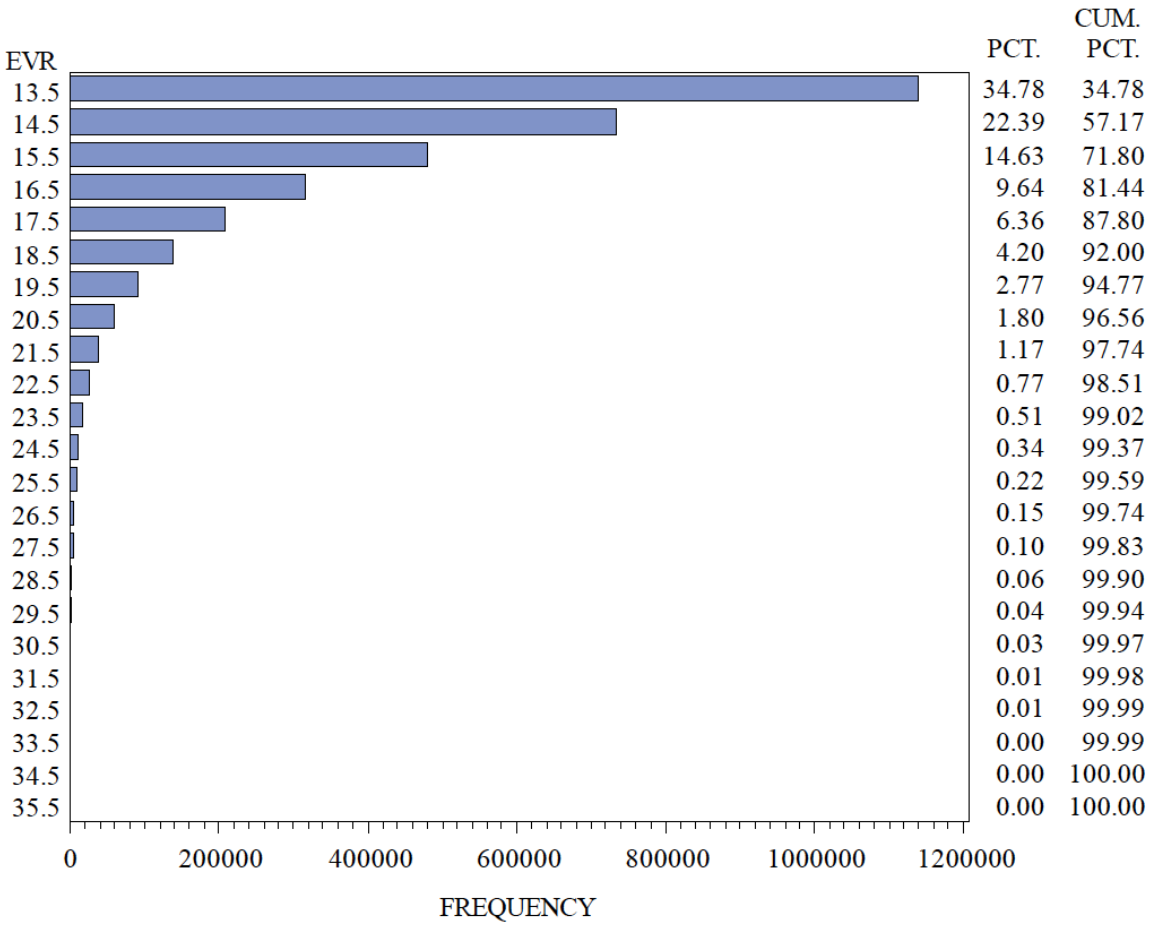

Figure 6. Distribution of Daily Maximum 8-hour Average EVR For Values of EVR≥13 (L/min-m^2)(midpoints on vertical axis)(Altanta 2006 base case, ages 18-35).

The binning of EVRs for use as moderate or greater exercise is a policy choice that EPA made first in 1996. It would be straightforward to evaluate the sensitivity to that choice in the final HREA and thereby evaluate the extent of bias in the current analysis. This should be done for the final documents. The impact of this one factor is probably greater than the differences between the alternative standards evaluated in the HREA and PA.

The third way the counts of benchmark exposures are biased high relates to the fact that human ozone exposures near a monitor are lower than the monitor measures. The 2006 Criteria Document acknowledged that ozone exposure is lower at "breathing" height compared to "measurement" height (3-15 meters). For example, Wisbeth et al. (1996)[58] measured the increment between ozone at 2 and 10 meters and reported an average 13

[57] HREA, supra note 1, at 6-34.
[58] A. Wisbeth, G. Meiners, T. Johnson, and W. Ollison *"Effect of monitor probe height on measured ozone concentration,"* Paper No. 96-RA111.02, presented at the 89th Annual Meeting of the Air & Waste Management Association, Nashville, TN, June 1996.

percent difference. In addition to the height differential, ozone monitors are also sited in open areas removed from sources so as to capture the highest ozone concentrations expected in an area. Since downwind sites are usually the design value sites, they will dominate the upper tail of the ozone distribution and yet may not reflect the overall outdoor exposures in the vicinity of the site. If people spend time outdoors in closer proximity to streets or in areas with more surface area (buildings, etc.) to quench ozone, their exposures will be below that measured at the monitor. The APEX model assumes that whatever ozone is interpolated from the monitor measurement is the actual ozone exposure in the outdoors microenvironment. The 2007 Langstaff Memorandum acknowledged the issue of vertical variation in ozone but indicated that the Agency did not plan to address it due to a lack of data. This vertical difference was corrected in the vegetation risk assessment in the previous review but not in the human risk assessment. In the vegetation risk, the metric summing concentrations of 60 ppb and higher was halved with a 10 percent vertical correction.[59] By analogy, a vertical correction in the human risk assessment would likely halve the number of human exposures of concern at ground level. Because this effect would correct a bias in the exposure calculations, it is particularly important that the HREA include a calculation of the sensitivity to this bias in the final document.

The HREA does acknowledge

> Differences between ground-level (0-3 meters) and building rooftop sited (25 meters) monitor concentrations can be significant. Most importantly, use of higher elevation monitors would tend to overestimate ground-level exposures (i.e., persons outdoors).[60]

and that "Given judged impact to exposure, additional characterization is possibly warranted."

1. Summary

Given the Agency acknowledgement that the three concerns raised in this section are legitimate and that each would decrease the benchmark headcounts substantially, it is incumbent on the Agency to carry out the requested sensitivity analyses. The results should be included in Chapter 5, considered in the analyses in Chapter 6, and discussed in the synthesis Chapter 9 and in the Executive Summary, and carried over to the PA so that the readers of the documents, including the policy and decision makers in the Agency, are aware of the biases and uncertainty in the clinical risk assessment.

To better put the results of the clinical assessment into perspective with regard to public health, the percent of person-days with exposures above the benchmarks should be presented in as much detail as the percent of persons with one or more exposure above the benchmarks. Finally, all the benchmark results are presented without any error bars, leaving the impression that we know these results with great precision. This is also

[59] 2007 SP, supra note 7, at pp. 7-46 and 7-47.
[60] HREA, supra note 1, at p.5-59.

misleading to the decision maker. A better visual presentation of the uncertainty in the estimates in needed.

B. Characterization of health risks based on clinical studies -- estimates of FEV1 decrements

The second listed goal in the HREA is to provide estimates of the number of people with various decrements in forced expiratory volume in one second (FEV1). AIR agrees with EPA that estimates of risk based on results of human controlled human exposure studies are valuable because there is clear evidence from these studies that there is a causal relationship between exposures to O_3 over multiple hours and reductions (or decrements) in the performance of lung function tests at moderate to severe levels of exertion.

While the calculations presented in Chapter 6 are necessary they are not sufficient to estimate either the risk of adverse effects or the impact on public health since they do not include estimates of lung function decrements accompanied by respiratory symptoms, as the American Thoracic Society Guidelines recommend.[61]

Chapter 6 indicates that the population risk estimates for lung function decrements (e.g., $\geq 10\%$, $\geq 15\%$, and $\geq 20\%$ reduction in FEV1) are estimates of the expected number of people who will experience that lung function decrement in a year, the number of times that people experience repeated occurrences of given lung function decrements, and the number of occurrences (person-days) of the given lung function decrement.

Chapter 6 reports the results of two approaches to estimate FEV1 decrement risk. The first uses probabilistic exposure-response (E-R) functions similar to the risk assessment in the prior review. These functions were applied to the APEX estimated population distribution of 8-hour maximum exposures for persons at or above moderate exertion ($\geq$ 13 L/min-m^2 body surface area) to estimate the number of persons expected to experience lung function decrements. The second approach, based on the McDonnell-Stewart-Smith (MSS) FEV1 model,[62] uses the time-series of O_3 exposure and corresponding ventilation rates for each APEX simulated individual to estimate their personal time-series of FEV1 reductions, selecting the daily maximum reduction for each person.

1. Comments on the E-R approach

The probabilistic E-R function is shown in Figure 6-6 and the results of the analysis are given in Table 6-7. As shown in Appendix 6C, the risks of FEV1 decrements based on the population exposure-response (E-R) function in the HREA are similar to those estimated in the last review with the E-R methodology. As shown in Table 6C-1, the percent of asthmatic school-age children with at least one FEV1 decrement ≥ 10 %

---

[61] "What Constitutes an Adverse Health Effect of Air Pollution? " Official Statement of the American Thoracic Society Adopted by the ATS Board of Directors, July 1999, *Am. J. Respir. Crit. Care Med*., 161, 665-673 (2000).

[62] W. McDonnell, P. Stewart, M. Smith, C. Kim, and E. Schelegle, "Prediction of lung function response for populations exposed to a wide range of ozone conditions." *Inhal. Toxicol.*, 24(10), 619-633. (2012).

ranges from about 2 to 7 percent at the current ozone standard. Similarly, the percent of all school age children with at least one FEV1 ≥ 15 % ranges from less than 1 to 2 percent at the current ozone standard. The responses in these analyses are calculated down to exposure concentrations of zero.

Since the output of the APEX model provides the exposure input for the E-R approach, the biases indentified by AIR and acknowledged by the Agency with respect to the benchmark headcounts translate directly into biases and overestimates of the FEV1 decrements with the E-R method. Thus, the sensitivity to these biases should also be carried over and evaluated in Chapter 6. In addition, the results for person-days over the ozone season as well as persons with one or more decrement should be presented for each scenario evaluated in Chapter 6. The results for person-days of occurrences should be presented as a percent of total person-days.

In comments on the first draft REA, AIR included a comparison of the two metrics, percent persons with one or more occurrence in the ozone season and percent of person-days over the ozone season, based on the Denver 2006 base case noted above for the benchmark comparison. It is shown in Table 6 and indicates that the portion of person days with various FEV1 decrements is well below 0.1 % even for a Denver base case where the ozone design value was 90 ppb as compared to the current 75 ppb standard.

Table 6 – Denver 2006 Base Case

Percent Persons				**Percent Person-Days**			
8-h>60	FEV>10	FEV>15	FEV>20	8-h>60	FEV>10	FEV>15	FEV>20
30	7.5	2.5	0.6	0.27	0.068	0.023	0.005

Thus, the APEX model coupled with the E-R model, even with its bias to over-predict ozone exposures and FEV1 decrements, predicts that the current standard is very protective. Only a small percentage of subjects will be exposed to the benchmark levels more than once per year, and only a small portion of those will experience FEV1 decrements that, in themselves, are small, transient, and will not interfere with daily living. The majority of the population, including school age children, will not experience even these mild effects at all during the year and, for those that do, they are protected almost all the time. Thus, based on the now extensive body of human clinical studies, the risk to public health at the current ozone standard is minimal.

2. Comments on the MSS approach

The MSS model results are given greater weight in the HREA than the E-R results. Although the model is capable of estimating individual responses, the analysis relies on APEX to generate the time-series of ozone exposure and corresponding ventilation rates for each simulated individual. That data is used to estimate the personal time-series of FEV1 reductions for each individual using the MSS model. Finally, the daily maximum reduction for each person is used as the output metric of choice. Since the MSS approach uses APEX, the concerns for biases in the ozone exposures and

ventilation rate extremes documented above apply to the output of the MSS model. Therefore, any sensitivity analyses carried out with APEX, or corrections to APEX should be carried over to the FEV1 analyses

The MSS data presented in Chapter 6 focuses on the percent of subjects experiencing one or more FEV1 decrements an ozone season. This statistic is not particularly informative as it relates to public health. As for the results in Chapter 5, it would be more reflective of the risk to public health to present the FEV1 decrement data as a portion or percent of the total person-days in the particular city and year for each base case and alternative standard. While there are data on the counts of person-days for each city, year, and alternative standard in Appendix 6B at pages B-17 to B-31, based on the MSS approach, the data presented in this manner are misleading and insufficient as regards public health.

First, the counts are presented with an undue precision, with up to 7 significant figures. Second, the counts, by themselves, are not a valid measure of either the chance of an individual's risk or the total group risk. One needs to consider the size of the population being simulated and the number of days being simulated to have perspective on what the counts may mean for public health. To interpret the counts one needs to refer to the data in Table 5-1 of the HREA to see the number of total persons, the number of school age children, and the length of the ozone season simulated. For example, the simulations of school age children in Detroit and Houston each consider about 1,000,000 children, but the ozone season in Houston is the total year while in Detroit it is six months. To evaluate the risk based on these counts, the counts need to be normalized by the total person-days in the simulation. When this is done the fraction (or percent) of person-days with various FEV1 decrements can be calculated. To continue the example, the counts of FEV1 $\geq$ 15 % for school age children are similar for Detroit and Houston for attaining the current standard using the 2008 base case, 120,439 and 130,076 respectively. However, when normalized the fraction (and percent) of total person-days is 0.00066 (0.066 %) in Detroit and 0.000367 (0.037 %) in Houston. Interestingly there are two estimates for 2008 based on the two design value periods (2006-2008) and (2008-2010) and they are 0.04 % and 0.066 % for Detroit and 0.04 % and 0.02 % for Houston. Instead of providing counts of person-days in the Appendix in an uninformative form, the HREA and PA should provide the estimates of percent of person-days along with the percent of persons with one or more event in the main documents.

Third, even the presentation of both metrics in terms of percent is insufficient as regards the protection of public health. It is also important to discuss how the FEV1 decrement results relate to public health. Isolated, small, transient and reversible FEV1 decrements without symptoms have not been considered adverse in prior reviews for either asthmatics or normal subjects.[63] In addition, the decrements need to be considered in relation to the known mechanism that is causing the decrements, the measurement error of the test, and the variability in subject responses. These issues will be discussed in the AIR comments on the PA.

[63] U.S. EPA. 1996. *Review of the National Ambient Air Quality Standards for Ozone Assessment of Scientific and Technical Information OAQPS Staff Paper*, EPA-452/R-96-007, June 1996.

The MSS model predicts more occurrences of various decrements than the E-R approach. As shown in Figures 6-9 and 6-10, the MSS model predicts FEV1 ≥ 10 % decrements at exposures as low as 10 to 20 ppb and predicts substantial decrements below 60 ppb. Also as shown in Table 6-10, almost half of the profiles with instances with FEV1 ≥ 10 % never experience 8-hour EVR ≥ 13. The HREA points out that the MSS model includes a threshold parameter that allows for modeling a delay in response until cumulative dose rate (taking into account decreases over time according to first order reaction kinetics) reaches a threshold value. McDonnell et al. 2012 found the inclusion of a threshold improved the model fit and the threshold model is the one used in the risk assessment. The HREA points out "the threshold is not a concentration threshold and does not preclude responses at low concentration exposures."[64]

The question arises as to why the MSS model predicts FEV1 decrements at low ozone concentrations and mild exercise rates even though the model includes consideration of a threshold. First, McDonald et al. acknowledge that the data from the individual lung function measurements are noisy. The model was developed from a dataset of 8477 lung function measurements during ozone exposure. There is also a dataset of 2948 measurements made during filtered air exposures. The fit of the individual model predictions versus the observations for the 8477 individual measurements during ozone exposure is shown in Figures 2a and 3a from McDonnell et al. 2012 as shown below in Figure 7. The noise in the individual response data is evident in these figures with the range of the data as the predictions approach zero being roughly between a 10% improvement in FEV1 to a 10% decrement. In fact, the HREA acknowledges that the model does not have good predictive ability for individuals, with $R^2 = 0.28$.[65]

[64] HREA, supra note 1, at p. 6-9.
[65] Ibid., at p. 6-38.

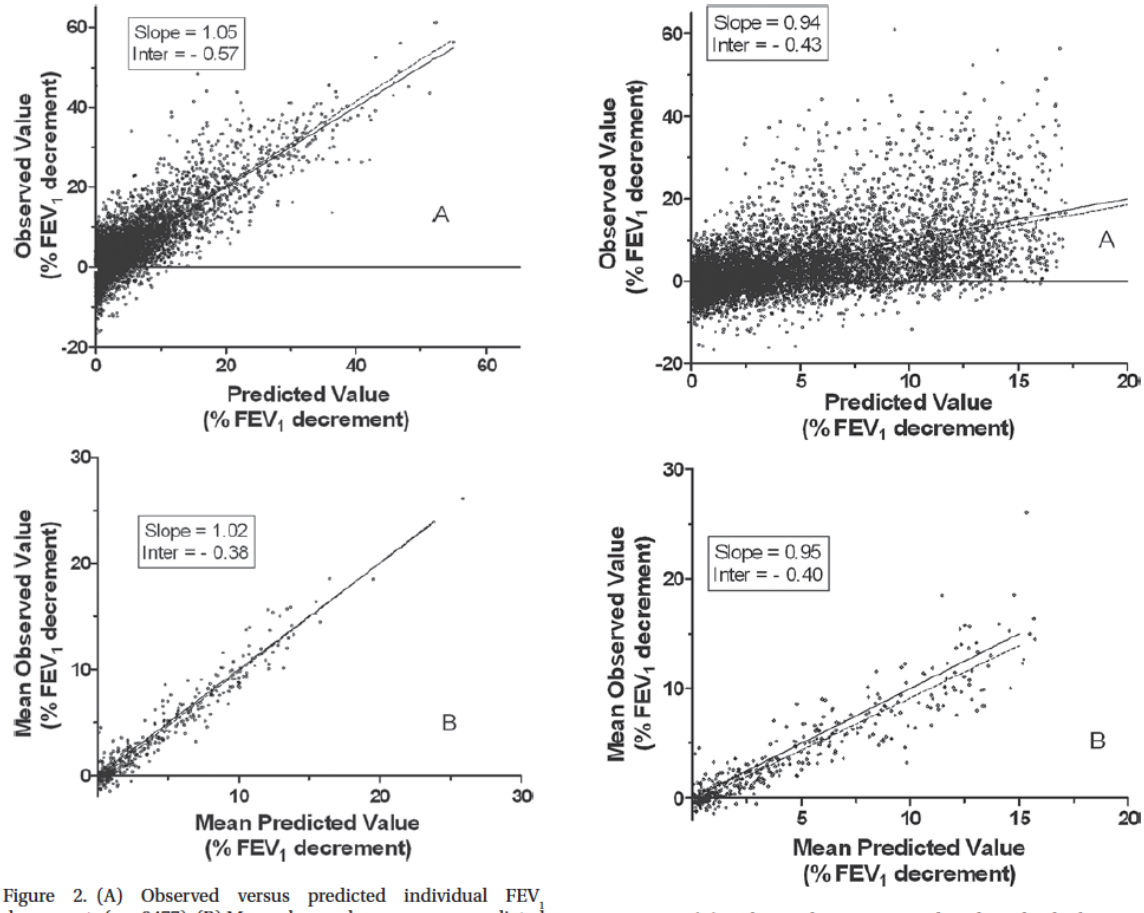

Figure 2. (A) Observed versus predicted individual FEV_1 decrements (n = 8477). (B) Mean observed versus mean predicted FEV_1 decrements. Means (across subjects) are calculated for each time point of each exposure condition of each study (n = 365 data points). Both figures use the empirical best linear unbiased predictor of U for each individual for calculation of predicted values. Solid line: identity; dashed line: regression.

Figure 3. (A) Observed versus predicted individual FEV_1 decrements (n = 8477). (B) Mean observed versus mean predicted FEV_1 decrements. Means are calculated for each time point of each exposure condition of each study (n = 365). Both figures use the population mean of U for calculation of predicted values. Solid line: identity; dashed line: regression.

Figure 7. Figures 2 and 3 from McDonnell et al. 2012.

Similarly when the model is used to predict the portion of responses greater than 10, 15 or 20% there is substantial variability in the individual predictions as shown in Figure 4 from McDonnell shown below in Figure 8. The substantial variability in the individual responses means that there will be predictions of both decrements and improvements in FEV1 in the model output. The largest decrements are counted in the EPA analysis so that it gives the appearance of potential risk at low exposures and ventilation rates when the group mean FEV1 changes are extremely small.

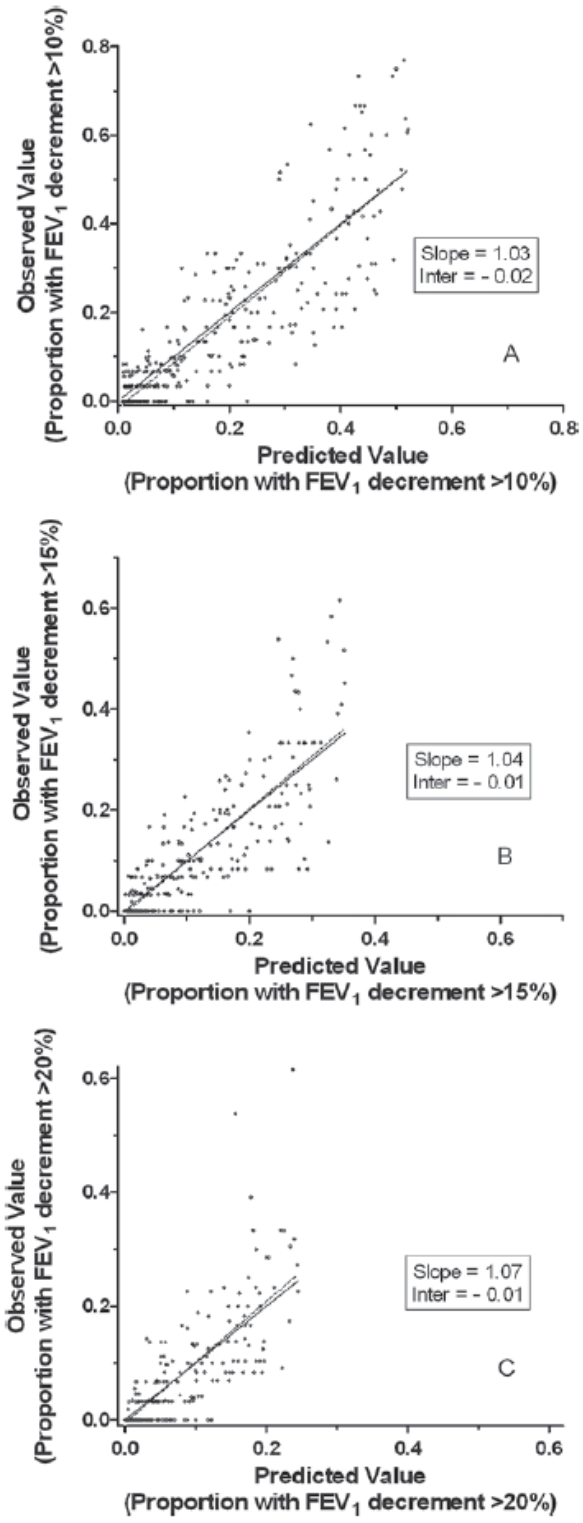

Figure 4. Observed versus predicted proportions of individuals experiencing an FEV_1 decrement greater than 10% (A), 15% (B), or 20% (C). Proportions are calculated for each time point of each exposure condition of each study (n = 365). Solid line: identity; dashed line: regression.

Figure 8. Figure 4 from McDonnell et al. 2012.

McDonnell et al. point out:

> All within-subject variability is currently lumped into a single term E as a result of limitations of the model fitting program. It is likely that some of the within-subject variability is due to true changes in responsiveness to ozone over time while much is simply noise.

The variability term E is discussed on page 6-41 where it is shown that the results are highly sensitive to this parameter. The model predictions are also highly sensitive to the parameter beta6 which is the power to which the ventilation rate is raised as noted on page 6-39. The noise in the data is also evident in the comparison shown in Appendix 6D where the filtered air exposure results in decrements and improvements of over 10 % in individual FEV1 changes for a group of 26 8-11 year old children in the McDonnell et al. 1985 study.

Since EPA has evaluated the sensitivity of the model predictions to a number of model parameters, it would be straightforward to evaluate the sensitivity of the output to the factors AIR and other public commenters have raised that will reduce the predicted exposures and FEV1 changes.

3. Summary

Both methods of estimating FEV1 decrements depend on the use of the APEX model, so the sensitivity to the biases in that model need to be carried over to the analyses and discussion in Chapter 6. Even with the biases, both the E-R and MSS method predict an extremely small percent of person-days with exposures that may result in FEV1 decrements. Both the percent of persons with one or more occurrence and the percent of person-days metrics should be presented and discussed in the final HREA.

Due to its high variability, the MSS model predicts some individual decrements at very low exposures and low exercise levels where the group mean decrements are extremely small. It is not clear whether these are real effects due to ozone or whether they are related to the noise in the underlying data. In addition, due to the prediction of effects well within background ozone, the final HEA should include estimates of the changes due to the change in U. S. man-made precursor emissions not the changes in total ozone. This is critical since it will provide an estimate of the risk reduction that is possible by reducing U. S. emissions.

C. Mortality Assessment

To develop a risk assessment, EPA had to select concentration-response (C-R) functions for each health endpoint. The criteria they used to select the individual C-R functions are:

- The study was peer-reviewed, evaluated in the O_3 ISA, and judged adequate by EPA staff for purposes of inclusion in the risk assessment.
- Preference for multicity studies.
- The study design is considered robust and scientifically defensible, particularly in relation to methods for covariate adjustment, including treatment of confounders, as well as treatment of effect modifiers.
- The study is not superseded by another study (e.g., if a later study is an extension or replication of a former study, the later study would effectively replace the former study), unless the earlier study has characteristics that are clearly preferable (e.g., inclusion of copollutants models, or use of a peak exposure metric of interest).[66]

Based on these criteria, EPA states:[67]

> For short-term exposure related mortality, our core analysis is based on application of C-R functions obtained from Smith et al., 2009[68] epidemiological study. In addition, we have completed an expanded array of sensitivity analyses which provide coverage for a number of modeling elements including: ...application of alternative C-R functions based on Zanobetti and Schwartz, 2008[69]....

They further state:

> Based on additional evaluation of the literature, we have substituted Smith et al., 2009 for Bell et al., 2004[70] as a source of Bayes-adjusted city-specific effect estimates to support modeling short-term O_3-attributable mortality. This decision reflects a number of factors. The Smith et al., 2009 study includes a wider range of simulations exploring sensitivity of the mortality effect to different model specifications including (a) regional versus national Bayes-based adjustment, (b) copollutants models considering PM_{10}, and (c) all-year versus O_3-season based

[66] U.S. EPA, supra note 12, at p. 7-18.
[67] Ibid, p. 7-4-7-5.
[68] Smith, R.L.; B. Xu and P. Switzer. 2009. "Reassessing the Relationship between O_3 and Short- term Mortality in U.S. Urban Communities." *Inhalation Toxicology*, 21: 37-61.
[69] Zanobetti, A; J. Schwartz. 2008. "Mortality Displacement in the Association of O_3 with Mortality: An Analysis of 48 Cities in the United States." *American Journal of Respiratory and Critical Care Medicine,* 177: 184-189.
[70] Bell, M. L.; A. McDermott; S. L. Zeger; J. M. Samet; F. Dominici. 2004. "O_3 and Short-term Mortality in 95 U.S. Urban Communities, 1987-2000." *JAMA,* 292: 2372-2378.

> estimates. This is contrasted with the Bell et al., 2004 study which does not provide this degree of model exploration.[71]

It is interesting that EPA selected the Smith et al. paper for their core analysis since the first sentence of the paper states: "The purpose of this paper is to reexamine the evidence of an association between ambient ozone and nonaccidental all-cause mortality, based in particular on a series of papers by Bell and co-authors that used the NMMAPS database." To accomplish this, they state: "We look extensively at alternative treatments of meteorology and co-pollutants, showing that there are confounding and effect modifier relationships that have been understated or overlooked in previous studies."

The relative risk value that EPA uses from Smith et al. to develop the C-R for non-accidental mortality was an increase of 0.32% ± 0.08 for a 10 ppb increase in MDA8 O_3. Smith et al. generated that number using a model that was identical to that used by Bell et al. to make sure they could first replicate Bell et al.'s result before conducting their sensitivity analyses. Since Bell et al. used 24-hour average O_3, Smith et al. first reproduced their result using the same model and then ran it a second time with MDA8 O_3 values to generate the relative risk value in terms of MDA8. Then Smith et al. conducted their sensitivity analyses by running many more alternative models and generated hundreds of different relative risk values that ranged from negative values to statistically significant positive values. Smith et al.'s analyses demonstrate that the 0.32% risk estimate is not robust to alternative model formulations. Smith et al. do not identify any one model as being the correct model as the point of their calculations was to show that different model specifications produce different answers. As a result, their conclusions include:

> - The basis for the national effect estimates published by Bell and others is questionable in the face of clear evidence that the [geographical] effect is not homogeneous.
>
> - Further, we believe that the heterogeneity and sensitivity of ozone effect estimates to a variety of covariates leaves open the issue of *whether or not ozone is causally related to mortality. Consequently, the question arises whether any particular ozone-mortality effect estimate can reliably be used to predict mortality reductions that would ensue from specific ozone reductions*[emphasis added].
>
> - There is clear evidence of a PM_{10} co-pollutant effect that has been understated or misinterpreted in previous publications.

[71] U.S. EPA, supra note 12, at p. 7-19.

- The nonlinear analysis shows that much of the evidence for an ozone-mortality relationship in fact comes from the low-ozone days, but human studies do not support an ozone effect at such low ozone levels. It is possible that the appearance of an association at low ozone levels may be due to the effect of co-pollutants, or an artifact caused by differences between personal and ambient exposure.

- There are other methodological issues that have not been discussed in this paper, but that could affect the results.

- In summary, it is our view that estimates of the association between ozone and mortality, based on time-series epidemiologic analyses of daily data from multiple cities, reveal important still-unexplained inconsistencies and show sensitivity to modeling choices and data selection. *These inconsistencies and sensitivities contribute to serious uncertainties when epidemiological results are used to discern the nature and magnitude of possible ozone-mortality relationships or are applied to risk assessment* [emphasis added].

In essence, EPA pulled one out of hundreds of risk estimates contained in the Smith et al. paper because it met their criteria and ignored many others. In addition, they make no mention of the conclusions that Smith et al. come to when all of the results of their analyses are considered in context.

Heterogeneity of results and the dependence of the results on model selection were also illustrated in the multi-continent APHENA study.[72] APHENA provides a particularly large data base and set of analyses with various statistical models that can be used to evaluate important questions concerning the ozone-mortality and ozone-hospital admissions associations. As documented in Appendix 1, the combined results of the large and comprehensive APHENA study are not consistent with ozone having a causal role in mortality or morbidity below the current standard. The authors of the HREA were clearly aware of APHENA because it is referenced in Chapter 7 of the HREA.

The strong regional differences in ozone-mortality associations that have now been identified should supersede the EPA assumption of a common national mortality health effect. In addition, the APHENA results, as discussed in detail in Appendix 1, indicate results that are mixed, inconsistent, and model-dependent.

The HREA acknowledge that there is heterogeneity in ozone-mortality associations. However, the heterogeneity is much wider than EPA acknowledges and includes many

[72] K. Katsouyanni and J. Samet. 2009. "Air Pollution and Health: A European and North American Approach", (APHENA), *HEI Report* 142, Oct. 2009.

cities with negative associations. The discussion of possible reasons for the heterogeneity in the HREA, and PA only discusses factors that could lead to varying degrees of positive association. In reality, especially for hospital admissions and mortality, the full pattern of associations in multi-city studies includes a substantial number of negative associations, a substantial number of null or near null associations, and a substantial number of positive associations. The full range of mortality associations as shown in Figures 6-28, 6-29, and 6-31 of the ISA varies between -5 % to +10% change in daily mortality for a 10 ppb increase in ozone.

In another example, in the Medina-Ramon et al., 2006 study of 36 U. S. cities the individual-city associations for COPD hospital admissions in the summer ranged from –30 % to +40 % for a 30 ppb increase in 8-hour ozone. The individual-city associations for pneumonia hospital admissions ranged from –15% to + 20% for a 30 ppb increase in 8-hour ozone. The combined associations for the two categories were positive in the warm season, but were negative in the cold season and not statistically significant over all year. By switching the baseline analysis for the Medina-Ramon study to the all-year result, the appropriate conclusion to be drawn is that the hypothesis that ozone has no effect on respiratory hospital admissions cannot be rejected.

It is important for policy makers to be given the full story concerning the range of associations in the literature and the spatial and temporal variations that have been reported. In addition, the role of publication bias inflating the magnitude of the perceived effect and the role of model selection uncertainty should be documented in the HREA. For example, it was noted in the prior review that variations in treatment of weather can change the results by a factor of 2 and that publication bias can inflate the perceived association by a factor of 3. There is also a Keatinge and Donaldson analysis (that has been ignored in the current review) indicating that previously overlooked weather factors can reduce the ozone mortality association by a factor of 10.[73]

To demonstrate the full range of associations, the HREA should include estimates of risk from the individual cities in the NMMAPS data that has been analyzed now by several investigators. It is fine to include Bayesian-adjusted results with both regional priors and national priors, but the unadjusted individual-city associations should also be shown to policymakers. Figure 4 in Smith et al. (2009) demonstrates the differences for MDA8 O_3 associations.

There is also strong evidence for unrecognized stochastic variability in associations within a given city. Ito (2003)[74] re-analyzed the 1220 separate air pollution mortality and morbidity associations that were included in the original Lippmann et al. (2000)[75] HEI study of Detroit. As shown in Ito's Figure 2, there was a wide range of negative and

[73] W. Keatinge and G. Donaldson, "Heat acclimatization and sunshine cause false indications of mortality due to ozone," *Environmental Research*, 100, 387-393 (2006).
[74] K. Ito. 2003. "Revised Analyses of Time-Series Studies of Air Pollution and Health," *HEI SpecialReport*, pp. 143-156.
[75] Lippmann M; Ito K; Nádas A; Burnett RT. 2000. Association of Particulate Matter Components with Daily Mortality and Morbidity in Urban Populations. Research Report 95. Health Effects Institute, Cambridge MA.

positive risks in Detroit when all pollutants, lags, and endpoints were considered. Ito showed in separate figures that the wide range of associations occurred for each pollutant. Although the focus in the original Lippmann study, as it is in almost all the published literature, was on the positive associations, Ito's plots shows that there are many negative associations in the data. Although there may be somewhat more positive associations than negative associations, there is so much noise or variability in the data, that identifying which positive associations may be real health effects and which are not is beyond the capability of current methods.

With regard to temporal variation, the NMMAPS analysis team showed that the combined ozone association was negative in the winter to the same degree that it was positive in the summer.[76] The same seasonal behavior is reported in the Medina-Ramon et al. (2006)[77] study of hospital admissions that is included in the HREA, with a negative combined association in winter and a positive combined association in summer. Since each of these studies is a large multi-city study, the temporal variation is robust. The HREA should present this information to policymakers. The implications of the full pattern of associations must be discussed in the PA.

As one demonstration of the uncertainty due to model selection, AIR compared the unadjusted individual-city ozone associations from the Zanobetti and Schwartz (2008) and Bell et al. (2004) for the cities the two studies have in common. The Zanobetti and Schwartz associations are shown in their Figure 1. The Bell et al. unadjusted associations are not given in the original paper but are shown in Figure 4 of Smith et al. (2009). As shown in Figure 9, there is little or no correspondence between the associations in individual cities in the two studies that EPA considers the best sources of data on this subject. Note that there are many negative associations in the data. For these unadjusted maximum likelihood estimates (MLE), there is one positive association and one negative association each for Baltimore, Boston, Philadelphia, and St. Louis. In addition, both MLEs are negative in Denver and both are essentially zero in Atlanta. By choosing the unadjusted MLEs for the baseline in the REA, a totally different picture concerning the likelihood of mortality due to ozone emerges in the 12 cities.

Another demonstration of model uncertainty is given in Figure 10 which compares the NMMAPS associations for individual cities that come from the 24-hour ozone associations at lag 1 from the 2003 revised analysis of time series data[78] with the ozone associations from the same cities using 8-hour ozone and the distributed lag model from Bell et al. (2004). Lag 1 was chosen for the comparison even though lag 0 had a somewhat higher combined association in the revised analysis because lag 0, in the case of ozone, runs afoul of the temporality requirement that the cause precede the effect. Since the peak ozone occurs in the late afternoon, the bulk of the mortality on a given day

[76] F. Dominici et al. 2003. "Revised Analyses of Time-Series Studies of Air Pollution and Health," *HEI Special Report*, pp. 9-24.
[77] Medina-Ramon, M.; A. Zanobetti and J. Schwartz. 2006. "The Effect of O_3 and PM_{10} on Hospital Admissions for Pneumonia and Chronic Obstructive Pulmonary Disease: A National Multicity Study." *American Journal of Epidemiology*, 163(6), 579-588.
[78] Ibid., at Figure 12.

occurs before the peak ozone exposure. Again the wide variation in associations for most cities is apparent in Figure 10.

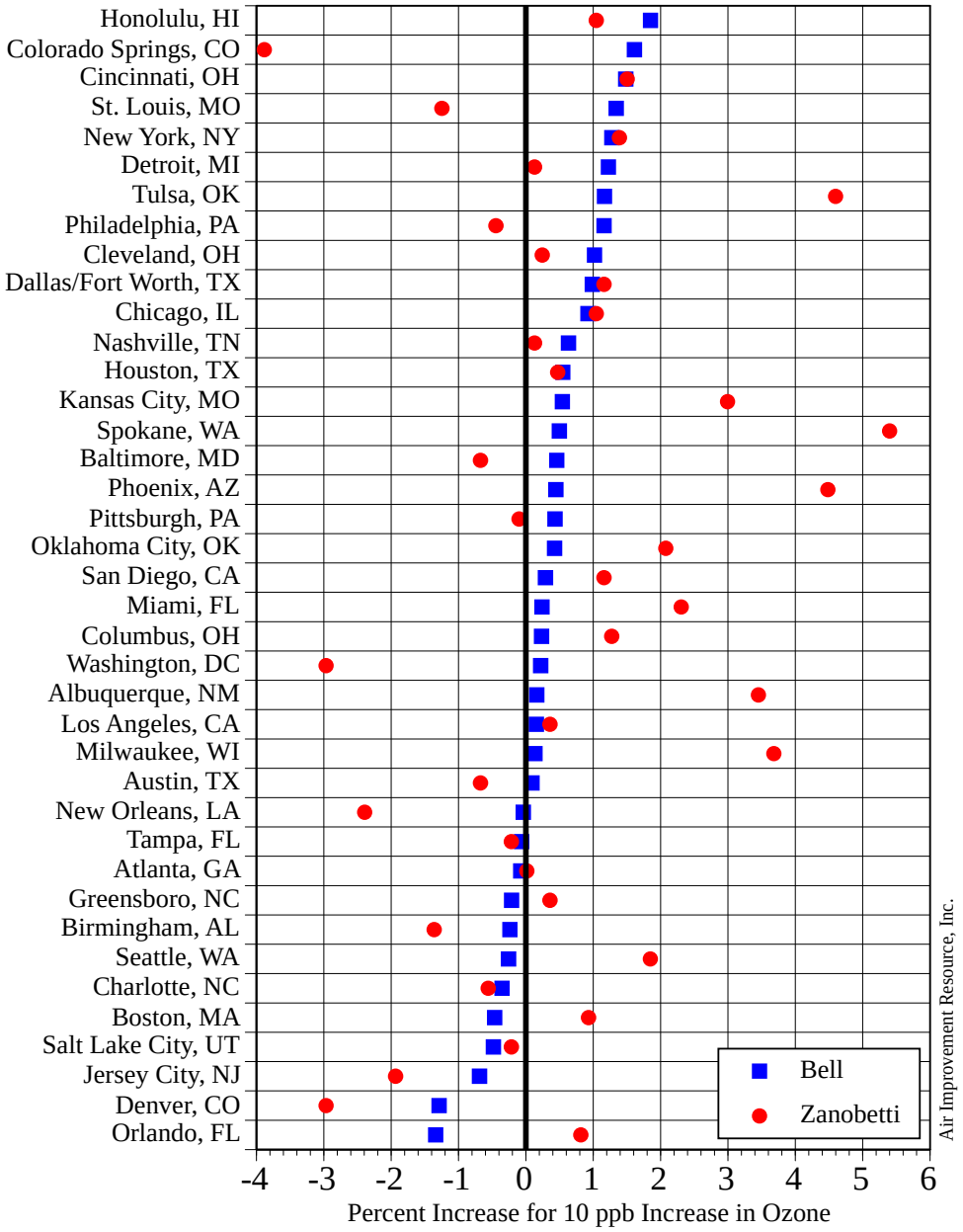

Figure 9: Comparison of unadjusted maximum likelihood estimates for mortality from Bell et al. (2004) and Zanobetti and Schwartz (2008).

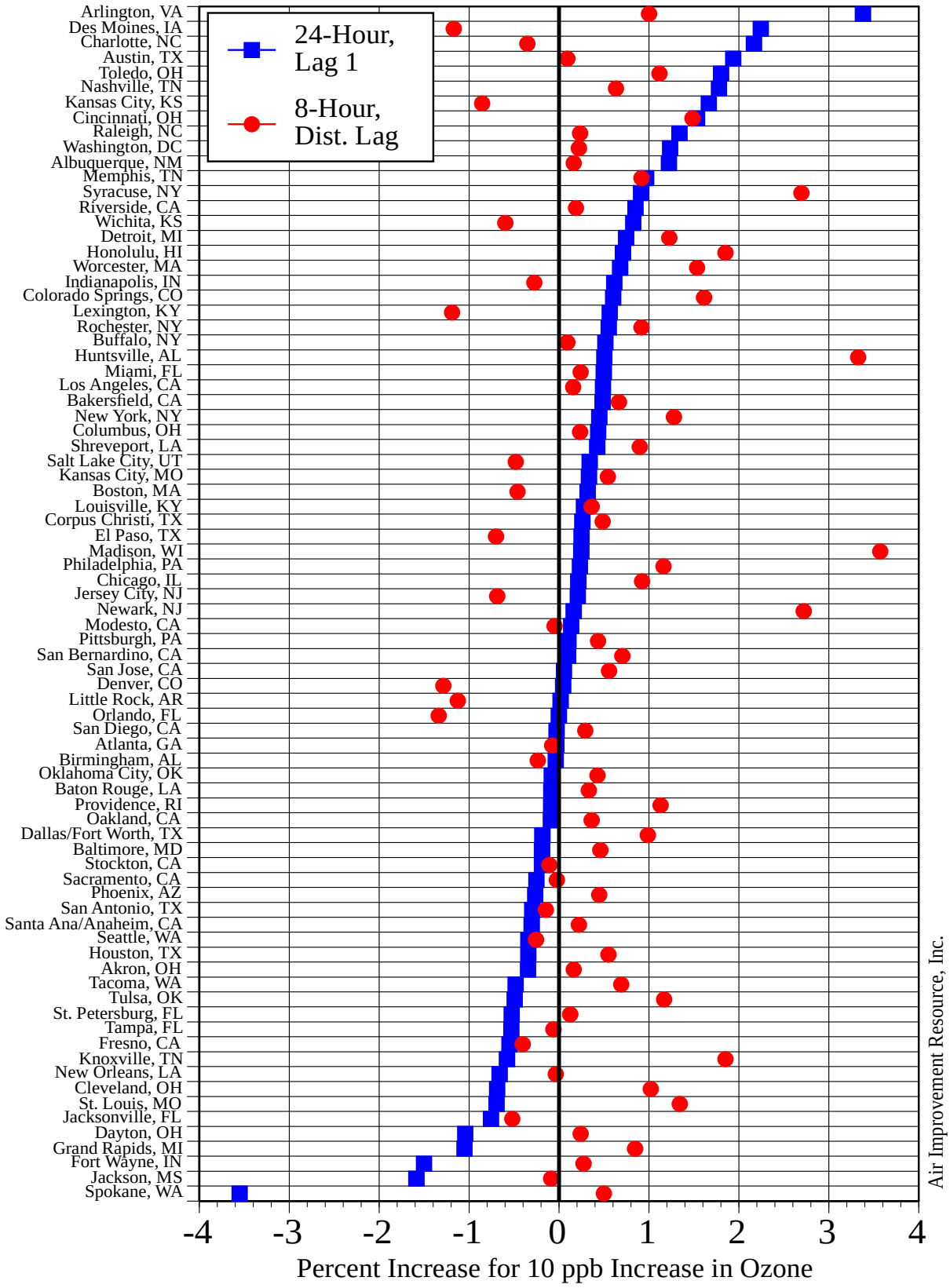

Figure 10: Maximum likelihood estimates for mortality from two NMMAPS analyses.

Instead of discussing the dose-plausibility of low ozone causing mortality, the HREA depends on selected epidemiology associations to estimate mortality effects. In discussing the shape of the concentration-response function, the ISA points out that combined mortality effects for ozone have been found at concentrations well below the current standard and cite a multi-city study where high ozone days have been excluded, Bell et al. (2006). However, there is a follow-on study by Bell et al. (2007)[79] that illuminates this issue. When Bell et al. (2007) restricted the analysis to days with low ozone, in order to see if the small combined association persisted, the range in individual-community associations widened. For example, when the data was restricted to days with ozone less than 20 ppb, the range in individual city mortality associations for a 10 ppb increase in ozone was from – 20 % to + 30 %. It is inconceivable that such low ozone exposures would be causing a dramatic increase in mortality in one city and protecting against mortality in another. With such wide variation, the interpretation of a small combined association as a health effect is highly questionable, especially in light of the fact that ozone indoors, where people spend about 90 % of their time is reduced about half or more by deposition to building surfaces.

With regard to chronic mortality, the HREA focuses on one positive study, Jerrett et al. (2009), as showing a chronic respiratory mortality signal for ozone. However, the respiratory mortality signal is present only for females in spite of the fact that males would be expected to receive higher ozone doses by being outside exercising more than females. In addition, the regional results reported by Jerrett et al. show no respiratory mortality effect in Southern California, the Northeast, or the Industrial Midwest, the regions of the country with the highest historic man-made ozone exposures. Moreover there are several other chronic mortality studies that do not report an ozone effect. The full body of studies is best characterized as inconsistent or inconclusive. Finally, the presence of a chronic respiratory mortality signal is not coherent with the lack of an acute respiratory mortality signal in the HEI multi-continent study. For these reasons, the evidence for a chronic ozone mortality effect is much weaker than indicated in the HREA and PA.

Rather than using EPA's preferred positive associations, AIR urges the Agency to explore the full range of associations in the literature. If this is done it will become apparent, as Koop and Tole pointed out in 2004:[80]

> Point estimates of the effect of numerous air pollutants all tend to be positive, albeit small. However, when model uncertainty is accounted for in the analysis, measures of uncertainty associated with these point estimates became very large. Indeed they became so large that the hypothesis that air pollution has no effect on mortality is not implausible. On the basis of these results, we recommend against the use of point estimates from time-series data to set regulatory standards for air

[79] M. L. Bell, J. Y. Kim, and F. Dominici, "Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time-series studies," *Environ Health Perspect*, 115, 1591-1595 (2007).

[80] G. Koop and L. Tole,"Measuring the Health Effects of Air Pollution: to What Extent Can We Really Say that People are Dying from Bad Air," *J. of Environmental Economic Management*, 47, 30-54. (2004).

pollution exposure.

The fact that the uncertainty due to model selection is much larger than the typical confidence limits on any given statistical association should be acknowledged in the HREA and PA and considered in the interpretation of the epidemiological data.

Because of all the issues with stochastic variability, publication bias, model selection uncertainty, confounding, etc. discussed above, time-series epidemiology of air pollution associations is a very blunt tool. CASAC raised this issue in a June 2006 letter to the Administrator noting that "[b]ecause results of time-series studies implicate all of the criteria pollutants, findings of mortality time-series studies do not seem to allow us to confidently attribute observed effects specifically to individual pollutants."[81] Further, due mainly to measurement error issues, CASAC questioned the likelihood of ozone itself causing mortality and noted the limitation that measurement error obscures thresholds in time-series studies, adding additional concerns about the utility of the time-series mortality estimates. More recently, Rhomberg et al. (2011)[82] have shown, as others have previously shown, that measurement error can give a false linear result. Although the Rhomberg et al. study of the impact of measurement error in environmental epidemiology was cited in public comments on the second draft ISA, it is still ignored by the Agency. CASAC's prior concerns and the Rhomberg et al. findings are consistent with points made by the Special Panel of the HEI Review Committee (Special Panel of the Health Review Committee, 2004)[83] that raised several cautions in interpreting the NMMAPS concentration-response results. They point out that measurement error could obscure any threshold that might exist, that city-specific concentration-response curves exhibited a variety of shapes, and that the use of Akaike Information Criterion may not be an appropriate criterion for choosing between models. The HEI Panel cautioned *that lack of evidence against a linear model should not be confused with evidence in favor of it* (emphasis added).

By assuming ozone mortality extends down to zero ozone and by using selected ozone-mortality associations from the literature, the HREA calculates a substantial burden of mortality even when man-made emissions are taken away. In fact, a whole chapter of the HREA, Chapter 8, is devoted to the exercise of estimating a national mortality burden from ozone. However, the full pattern of associations in the literature is not consistent with ozone causing mortality, the shape of the concentration-response is not known, and epidemiology studies cannot be used to identify a threshold because of exposure uncertainty. Consequently EPA's extrapolations of risk at low ozone concentrations in the HREA are not justified.

Given that the small positive results from time-series studies may reflect residual bias of

---

[81] CASAC Letter, EPA-CASAC-06-07, June 5, 2006, at 3 and 4.

[82] Lorenz R. Rhomberg, Juhi K. Chandalia, Christopher M. Long, and Julie E. Goodman, "Measurement error in environmental epidemiology and the shape of exposure-response curves," *Critical Reviews in Toxicology*, Sept. 2011, Vol. 41, No. 8; pp. 651-671. (doi: 10.3109/10408444.2011.563420).

[83] Special Panel of the Health Review Committee. Commentary. In: *The National Morbidity, Mortality, and Air Pollution Study Part III: Concentration-Response Curves and Threshold for the 20 Largest US Cities*, HEI Report 94, Part III, pp. 23-30 (2004).

the models due to weather, temporal or other unaccounted confounding factors, EPA cannot and should not draw conclusions on causality from these studies or use point estimates to set air quality standards.

III. Comments on the PA

A. Health Effects of Ozone and Their Public Health Significance

The discussion concerning ozone health effects and the public health significance of attaining the current ozone standard occurs in Chapter 3 of the PA. In particular, Section 3.1 discusses the evidence-based considerations, Section 3.2 discusses the risk-based considerations, Section 3.3 discusses CASAC input, and Section 3.4 discusses the adequacy of the current standard.

AIR has reviewed the draft PA as it relates to the primary NAAQS and concludes that it (1) overstates the nature and magnitude of ozone health effects and perceived risk to public health from current ozone levels, and (2) strains to make the case for inadequacy of the current ozone standard.

Chapter 3 overstates the consistency and coherence of the evidence. With regard to hospital admissions and mortality, the overall results of a large multi-continent Health Effects Institute (HEI) study do not support EPA's assumption of causal relationships between ozone and these endpoints. In particular with regard to respiratory mortality, EPA makes claims for consistent effects that are contradicted by the views of the original investigators and the HEI Review Committee. In addition, the issues of model selection uncertainty, confounding, and publication bias are ignored or downplayed in the Chapter. The final PA should address all these issues in the interpretation of the observational studies and their integration with the full range of ozone effects studies.

AIR comments focus first on the interpretation of the controlled human exposure studies since, as noted in the PA, they

> …provide clear and compelling evidence for an array of human health effects that are directly attributable to acute exposures to O_3 per se (i.e., as opposed to O_3 and other photochemical oxidants, for which O_3 is an indicator, or other co-occurring pollutants).[84]

In contrast to the controlled exposures studies, the epidemiologic studies that report associations with ozone are difficult to interpret because of stochastic variability, model selection uncertainty, publication bias, and potential confounding.

The introductory chapter in the PA indicates:

> In this draft PA, we consider the evidence from controlled human exposure studies in two ways. First, we consider the extent to which controlled human

[84] PA, supra note 2, at p. 1-21.

> exposure studies provide evidence for health effects following exposures to different O_3 concentrations, down to the lowest-observed-effects levels in those studies. Second, we use such studies to inform our evaluation of the extent to which we have confidence in health effect associations reported in epidemiologic studies down through lower ambient O_3 concentrations, where the likelihood and magnitude of O_3-attributable effects become increasingly uncertain.[85]

AIR comments, therefore, concentrate on these two considerations, what we can say about the health effects of ozone from the human clinical studies, and what the clinical studies can tell us about the likelihood of more serious effects like hospital admissions and mortality being caused by ozone.

1. Human Clinical Studies and Their Interpretation

As indicated in the ISA,[86] the controlled human exposure studies provide a strong and quantifiable body of information on the dose-response of effects of 1-to-3 hour and 6- to 8-hour exposures to ozone. The HREA notes that over 140 human clinical studies are referenced in the final ISA.[87] As the ozone concentration is increased from a filtered air control, the first effects, which are transient and reversible FEV1 decrements, are the body's reflexive reaction to the presence of an irritant gas unrelated to sensations of discomfort. Such effects occur after exposures to 80 ppb for 6 to 8 hours when the subjects are exercising at a rate that would be considered strenuous when carried out intermittently for an eight-hour period. Whether such effects occur at 60 or 70 ppb with this level of exercise has been highly controversial since the answer depends on how the baseline is evaluated, how the precision of the test is considered, how the day-to-day variability of a subject is evaluated, and how the data is statistically analyzed. During the previous review, the Adams (2006) study was the only study available at concentrations below 80 ppb. The Schelegle et al. (2009) and Kim et al. (2011) studies are now also available.[88] As noted in the PA these studies now all indicate very small group mean changes in FEV1 at 60 ppb, with an average response (adjusted for the response to filtered air) of 2.7 %. This small change in the performance of the test is of the same magnitude as the accuracy of repeat FEV1 measurements. Importantly respiratory symptoms were not affected by ozone exposure at the 60 ppb level.

With subjects at rest, the threshold for appearance of the first mild effects occurs at between 300 and 500 ppb. Thus, substantial exercise, which increases the dose of ozone inhaled, is necessary to elicit any effects near the current standard. In the studies with intermittent exercise, a clear dose-response is observed in FEV1 after a threshold dose is inhaled. The PA acknowledges that antioxidants within the airway lining fluid have been shown to prevent ozone-mediated cellular and tissue effects. The ISA acknowledges "The first line of defense against oxidative stress is antioxidants-rich ELF

---

[85] Ibid., at p. 1-22.
[86] U. S. Environmental Protection Agency, *Integrated Science Assessment for Ozone and Related Photochemical Oxidants*, EPA/600/R-10/076F, February, 2013, at p. 6-2.
[87] HREA, supra note 1, at p. 6-6.
[88] Literature referred to by author and date without a footnote are references included in the PA.

which scavenges free radicals and limits lipid peroxidation."[89] Therefore, only ozone exposures of sufficient duration and concentration will overwhelm the body's antioxidant defenses and allow oxidative damage to occur. The PA acknowledges[90] that a key event in the mode of action of ozone is the activation of neural reflexes that lead to involuntary truncation of inspiration which results in lung function decrements. The PA also points out that evidence is accumulating that secondary oxidation products are responsible for this effect. It is also relevant that children experience similar spirometric responses to young adults but have a lower incidence of accompanying symptoms and that older adults have reduced ozone-induced spirometric responses.

In addition to FEV1 decrements, there is substantial evidence that mild inflammatory processes occur in the lung that increase with increased dosage of ozone. The subjects in the human clinical studies also report respiratory symptoms, such as cough, shortness of breath, and pain on deep inspiration, that increase with the ozone dose. As with FEV1 decrements, there is clear evidence of a threshold in the inflammatory and symptom responses.

With regard to the inflammatory response, the most common indicator has been measurements of neutrophils obtained from the lungs of subjects by bronchoscopy at various time within the first 24-hours after exposure. A footnote in the PA explains

> Referred to as either neutrophils or polymorphonuclear neutrophils (or PMNs), these are the most abundant type of white blood cells in mammals. PMNs are recruited to the site of injury following trauma and are the hallmark of acute inflammation. The presence of PMNs in the lung has long been accepted as a hallmark of inflammation and is an important indicator that O3 causes inflammation in the lungs. Neutrophilic inflammation of tissues indicates activation of the innate immune system and requires a complex series of events, that then are normally followed by processes that clear the evidence of acute inflammation.[91]

The PA refers to a meta-analysis of 21 controlled human exposure studies (Mudway and Kelly, 2004) that involved O_3 exposures from 80 to 600 ppb, exposure durations from 1 to 6.6 hours, and from light to heavy exercise. While the PA indicates that Mudway and Kelly reported that PMN influx in healthy subjects is linearly associated with total ozone dose, the actual paper indicates that there is a threshold in the dose-response and one rationale for the study was that "Establishing these relationships is vital in determining threshold doses of ozone below which adverse responses are negligible in the healthy population."

With regard to symptoms, the PA refers to the McDonnell et al. (1999) study to indicate that symptoms increase with increasing ozone concentrations, duration of exposure and activity level. The data reported by McDonnell et al. clearly show a threshold

---

[89] ISA, supra note 79, at p. 6-24.
[90] PA, supra note 2, at p. 3A-1.
[91] PA, supra note 2, at 3-17.

phenomenon. For example, ozone exposures of 2 hours at rest with concentrations up to 300 ppb caused no symptoms.

For all three of the first identified effects of ozone exposure, threshold behavior is evident in the data. In each case, as the ozone dose increases either due to longer exposures or greater exercise, the first effects are very mild. As the dose increases, the effects become larger and likely apparent to the subject.

Another factor that needs to be included in the discussion is that there is heterogeneity of response among the subjects for these endpoints. There are now several studies showing that the most responsive subjects for one endpoint do not necessarily have the strongest responses for another endpoint. For example, McDonnell point out that it is quite common for a person exposed to a high concentration of ozone to have a large change in FEV1 while experiencing no or mild symptoms. The conclusion drawn by various authors who have compared FEV1, neutrophil, and symptom responses is that there are separate mechanisms or modes of action that are involved in the responses.

With this background as to what the nature of the first effects are as determined in controlled human exposures, the relevance of these effects to public health can be discussed.

a. The Public Health Significance of the First Effects of Ozone Are Not Adequately Discussed in the PA

The important question is not whether the small changes in the performance of lung function tests are statistically significant; the important question is their medical or public health significance. The PA does not adequately lay the groundwork for answering this question. Rather the PA strains to make the case that FEV1 decrements are "potentially adverse," "abnormal," or "clinically significant." The PA refers to several publications regarding guidelines for determining clinically meaningful FEV1 changes. Two of the references (ATS, 1991 and Pellegrino et al., 2005) discuss the use of lung function testing to evaluate various obstructive and restrictive disease states that result in changes in lung function. For example, the Pellegrino et al. (2005) review discusses lung function changes as they relate to progressing disease or the response of disease states to therapy. Pellegrino et al. do not discuss the clinical significance of the kind of transient, reversible changes caused by ozone. They do note, however, that statistical significance and clinical significance do not follow one another. They point out that two lung function measurements that are statistically indistinguishable may provide reassurance in a patient receiving therapy for a disease that is otherwise rapidly progressive. They note that the same tests may be very disappointing if one is treating a disorder that is expected to improve dramatically with the therapy prescribed. They also point out that a statistically significant change may be of no clinical importance to the patient.

The relevant reference is the American Thoracic Society guidelines regarding what constitutes adverse air pollution effects. The 1999 Guidelines indicate:[92]

[92] "What Constitutes an Adverse Health Effect of Air Pollution? ", Official Statement of the American

> The committee recommends that a small, transient loss of lung function, by itself should not automatically be designated as adverse. In drawing the distinction between adverse and nonadverse reversible effects, this committee recommended that reversible loss of lung function in combination with the presence of symptoms should be considered adverse.

Note that the ATS guidelines do not specify a specific cut-point for FEV1 decrements. Because of the linking of functional changes with symptoms, the PA should discuss the symptom results for the human clinical studies along with the FEV1 results to provide appropriate information to the reader. In the Adams (2006) study, the total mean symptom scores were only 2-4 units at 40 and 60 ppb out of a possible total score of 160. Adams indicated that the differences in the symptoms between the 40 and 60 ppb exposures and the filtered air control were not statistically significant. Kim et al. (2011) also indicate that the symptom scores were not different between ozone and clean air. Schelegle et al. (2009) indicate that the symptom scores were increased at 70 and 80 ppb but not at 60 ppb, with the symptom score at 70 indicating very mild symptoms.

Thus, according to the ATS guidelines, the functional changes at 60 ppb would not be considered as adverse. The PA should expand on the clinical and public health relevance of the functional effects. The basic nature and extent of functional effects has not changed since the 1997 and 2008 reviews. There is now data between 0 and 80 ppb, but the assumption made in 1997 was that functional effects, albeit small, do occur below 80 ppb. In the 1997 review, single incidences of the effects at 80 ppb (for either healthy or asthmatic subjects) were not considered to be adverse by CASAC and EPA staff. Nothing in the body of controlled studies has changed to alter that view. If anything, the growing evidence that the functional effects are an involuntary inhibition of maximal inspiration caused by activation of neural reflexes should reduce the concern over isolated transient, reversible lung function decrements.

For example, in the 1997 review, single incidences of the FEV1 effects at 80 ppb (for either healthy or asthmatic subjects) were not considered to be adverse by CASAC and EPA staff. The 1996 Staff Paper included extensive discussion of how to interpret the clinical results in terms of public health.[93] Large functional changes, > 20 % FEV1 decrements, and severe symptomatic responses were indicated as clearly adverse. Moderate functional changes and moderate symptoms were discussed in relation to interference with normal activity for both healthy and asthmatic individuals. For asthmatics, the Agency and CASAC concluded that moderate responses, when repeated, should be considered adverse. After considerable discussion there was consensus on CASAC that single, acute moderate functional responses should <u>not</u> be considered adverse for healthy individuals. Rather the staff indicated that the number of exposures

---

Thoracic Society Adopted by the ATS Board of Directors, July 1999, *Am. J. Respir. Crit. Care Med*., 161, 665-673, 2000.

[93] U. S. Environmental Protection Agency, Review of the National Ambient Air Quality Standards for Ozone: Assessment of the Scientific and Technical Information, OAQPS Staff Paper, EPA-452/R-96-007, June 1996 at pp. 62-72.

resulting in moderate responses should be considered a factor in determining adversity for healthy individuals. The category of moderate functional changes without symptoms or with minimal symptoms, the effects that occur below the current standard in some individuals, should be of even less of an issue with regard to protecting the public health.

Instead of depending on the ATS guidelines for adverse health effects of air pollution, the PA also refers to the 2000 guidelines for methacholine and exercise challenge testing. In contrast to normal individuals where exercise usually results in a small increase in FEV1, exercise induces airway narrowing in the majority of patients with asthma. The guidelines relate to testing of potential asthmatics to determine whether exercise induces bronchoconstriction. The PA indicates that greater than a 10 % change in FEV1 is considered abnormal by those guidelines and uses that fact to infer that a greater than 10 % change in FEV1 that is ozone-induced is clinically important. This is not warranted since the 10 % change due to exercise induced bronchoconstriction is actually a narrowing of the airways and a 10 % ozone-induced change is an inhibition of maximal inspiration during the test due to ozone's effect on neural receptors, as first proposed by Hazucha et al. (1989) and as documented in the ISA. This difference is very important and should be acknowledged in the PA.

The PA uses the 10 % response cutoff to claim effects at exposures to 60 ppb, referencing the ISA:[94]

> Though group mean decrements are biologically small and generally do not attain statistical significance, a considerable fraction of exposed individuals experience clinically meaningful decrements in lung function.

There are several problems with this claim. First, as noted by public comments on the first draft ISA,[95] the studies of exposure to 60 ppb with exercise were not designed to assess individual responses. To determine whether lung function changes for a given individual were due to ozone, an acceptable study design would include repeat measurements for each individual and utilize a scientifically acceptable statistical test on the data for each individual. Second, the individual data that is available demonstrates sufficient variability (with examples of individual responder's responses at 60 ppb greater than at 80 ppb) such that EPA's assumption that all FEV1 changes are due to the ozone exposure cannot be supported. Within-subject variability needs to be understood and accounted for before the ozone-induced effect can be determined. Third, the sample size is too small to generalize the results. Fourth, as noted above, EPA's claim regarding the 10 % response cut-off is not soundly based. Instead, the PA should discuss the public health significance of ozone-induced FEV1 changes at 60 ppb with exercise in light of the neural reflexive mechanism and the lack of any respiratory symptoms.

The immune system responses noted by EPA as the first indications of "inflammation" are physiological processes that occur in all living organisms under the stimuli of daily life. The first reported changes are small and reversible and well within the range of

[94] ISA, supra note 79, at p. 6-20.
[95] Goodman Rpt., supra note 6.

physiological variability. They fall into the category of biochemical markers that the American Thoracic Society indicates do not necessarily imply adversity. For example, the PA indicates:

> The ATS also concluded that elevations of biomarkers such as cell types, cytokines and reactive oxygen species may signal risk for ongoing injury and more serious effects or may simply represent transient responses, illustrating the lack of clear boundaries that separate adverse from nonadverse events.[96]

The 2000 review by Mudway and Kelly[97] notes that for neutrophils transiting into the lung - one of the earliest of these responses - it is not clear if the response should be considered beneficial (functioning to clear necrotic cells) or detrimental (leading to an active inflammation with tissue injury). The 2006 Criteria Document noted that generally, "the initiation of inflammation is an important component of the defense process; however, its persistence and/or its repeated occurrence can result in adverse health effects."

Since there is a threshold for even the first indications of an inflammatory response as for FEV1 decrements, the likelihood of persistent or repeated lung function decrements or inflammation is very small. For example, the typical ambient concentrations of ozone in recent years are quite low compared to the thresholds for the first physiological effects as determined from controlled exposure studies. The ISA indicates that "the median 24-h avg, 8-h daily max, and 1-h daily max O_3 concentrations across all US sites reporting data to AQS between 2007 and 2009 were 29, 40, and 44 ppb, respectively."[98]

In addition to the typically low ambient concentrations, the data on indoor/outdoor ratios and personal exposures in Section 4.3 of the ISA clearly show that personal exposures are only a fraction of the levels measured at ambient monitors. Typically, personal exposures average a quarter or less of the ambient measurements, even for school children that spend an average of two hours per day outdoors. Even for a group of camp counselors, the personal exposures averaged less than half of the ambient measurements. The first draft ISA concluded that "Another important finding is that the magnitude of personal exposures is smaller than concentrations reported at fixed-site monitors due to time spent indoors and the low indoor penetration of O_3."[99] The second draft indicates "personal-ambient ratios are typically 0.1- 0.3"[100] although individuals spending substantial time outdoors such as outdoor workers may experience higher ratios. When one considers that the median 8-h daily maximum ozone concentration across the country is 40 ppb and the personal exposures of the population are typically only a small fraction of the monitored concentration, it is clear that the day-in day-out exposures of the population are typically way below the threshold for the first physiological effects. This

---

[96] PA, supra note 2, at p. 3-55.
[97] I. Mudway and F. Kelly, "Ozone and the Lung: A Sensitive Issue," *Mol. Aspect. Med.*, 21, 1-48 (2000).
[98] ISA, supra note 79, at pp. 3-117.
[99] ISA, supra note 8, at pp. 4-8.
[100] ISA, supra note 9, at pp. 4-42.

is an important consideration in evaluating the public health significance of the effects identified in the human clinical studies. The low personal exposures to ozone even at ambient ozone levels that exceed the current standard, provide a large margin of safety from the first effects identified in controlled human studies for the vast bulk of the population as they go about their daily activities.

In addition, the possibility that peak exposures result in effects also needs to be considered. The 99th percentile of the 8-h daily maxima is 80 ppb and the 4th highest daily 8-h maxima now range from about 65 to 85 as shown in Table 3-4 and Figure 3-44 of the ISA, with many sites still exceeding the current 75 ppb standard. Although these peak concentrations overlap with the thresholds for the first effects, it should be borne in mind that a subject has to be outside, exercising at the time and place of high ozone for there to be an exposure that could cause an effect. Thus, the results of the clinical studies cannot be used directly to claim effects below the current standard.

Rather, they must be used to evaluate the risk by mapping the results onto realistic exposure/activity patterns. In order to calculate the risk, all the relevant factors need to be taken into account, with the role of exercise being particularly important. Indeed, this is what the APEX model attempts to do. As detailed in the AIR comments on the HREA, the APEX-based results for exposures of concern and FEV1 decrements are biased high. Nevertheless they show that the overall risk to public health is minimal.

The final PA should show the results for percent person-days of occurrences to put the risks in perspective. As discussed in the AIR comments on the HREA, the portion of total person-day occurrences of the various benchmark exposures and FEV1 decrements are extremely small, the order of a fraction of a percent. This is an indication of very small potential impact on public health of ozone at, or even slightly above, the current standard.

Isolated occurrences of transient, reversible FEV1 decrements have never been considered adverse in prior reviews. In fact the PA acknowledges:

> Although some experts would judge single occurrences of moderate responses to be a ‘‘nuisance,’’ especially for healthy individuals, a more general consensus view of the adversity of such moderate responses emerges as the frequency of occurrence increases.[101]

Therefore, the percent of person-day occurrences is of importance in judging the likelihood of public health impacts.

The presentation and discussion of the exposures of concern is not a measure of adversity since it is only a measure of the exposure, not the response. The presentation and discussion of FEV1 decrements is also, by itself, not a measure of adversity since it doesn’t include consideration of symptoms.

[101] PA, supra note 2, at p. 3-57.

The concern over repeated exposures is that they could interfere with normal activity, lead to increased medication use, or set the stage for more serious illness. However, the PA also indicates that

> For active healthy people, including children, moderate levels of functional responses (e.g., FEV_1 decrements of $\geq 10\%$ but $< 20\%$, lasting 4 to 24 hours) and/or moderate symptomatic responses (e.g., frequent spontaneous cough, marked discomfort on exercise or deep breath, lasting 4 to 24 hours) would likely interfere with normal activity for relatively few sensitive individuals.[102]

Concentrations at or below the current standard even with prolonged exercise have not been shown to result in moderate or stronger symptoms such as frequent spontaneous cough or marked discomfort on exercise or deep breath. This is another indication that the counts of FEV1 decrements at or below the current standard overestimate the risk to public health. Even exposures at much higher concentrations in the many human clinical studies, including exposures that did result in strong symptoms have proven to be remarkably safe for the subjects. Rom et al. (2013) point out that the human clinical studies for ozone, that include exposures up to 600 ppb,

> …have been remarkably safe; even exposure of members of sensitive subgroups, including individuals with asthma and individuals with atherosclerosis, appears so far to have a most minimal risk of severe adverse effects requiring medical intervention.[103]

Another potential concern raised in the PA is that a 10% decrement in FEV1 can lead to respiratory symptoms, especially in individuals with pre-existing pulmonary or cardiac disease, noting that people with chronic obstructive pulmonary disease (COPD) have decreased ventilatory reserve. This is a speculative concern. As CASAC has indicated, for ethical reasons, controlled exposure studies involve effects that are relatively mild and reversible, including changes in pulmonary function and increased evidence of inflammatory changes. Controlled studies of groups of asthmatics and COPD patients have been conducted with intermittent exercise at substantially higher ozone exposures than the current standard, resulting in group-mean FEV1 decrements as high as 20 to 25 %, suggesting that such effects are relatively mild with regard to clinical or public health significance.[104]

The HREA shows that the responsiveness to ozone decrements declines with age, with the counts of decrements for ages between 36 and 55 being dramatically reduced compared to the younger ages. Also the counts for greater than 55 are not shown but are described by the Agency as minimal. Since those with pre-existing pulmonary or cardiac disease are less likely to exercise vigorously than the normal population, the HREA

---

[102] Ibid., at p. 3-56.
[103] W. M. Rom, H. Boushey, and A. Caplan. (2013) Experimental Human Exposure to Air Pollutants Is Essential to Understand Adverse Health Effects. American Journal of Respiratory Cell and Molecular Biology **49**:5, 691-696.
[104] See Table AX6-3 in Vol. II of 2006 Ozone CD.

results show that any such risk at the current standard is minimal.

When the extremely small portion of person-days of occurrence is considered along with the small portion of persons with one or more occurrence in a year, it is clear that the current standard is very protective of public health.

b. What Do the Human Clinical Studies Tell us About the Likelihood of the Epidemiological Associations Representing Real Health Effects?

There is a major disconnect between the assumption that ozone is causing premature mortality with a concentration-response down to zero and the results of over 140 human clinical studies evaluated in the ISA. The human clinical studies demonstrate threshold behavior and the first effects above the threshold are mild, transient, and reversible. The HREA documents that such effects are rare at the current standard for children and young adults and even less for older individuals.

In contrast, the mortality analysis predicts that the bulk of the acute mortality occurs on days with ambient concentrations below 60 ppb.[105] Similarly the large burden of predicted chronic mortality occurs on days below 60 ppb. When the low ambient exposures and even lower personal exposures to ozone are considered, it is apparent that the mortality analyses in the HREA and PA are based on the assumption that daily personal exposures of the order of 10 to 20 ppb are causing premature mortality. There is nothing in the human clinical studies or the toxicological studies that show how this could possibly occur.

The PA is silent on the disconnect except to dismiss the consideration of human clinical studies in a footnote on page 7-30 of the HREA noting:

> The clinical studies focus on relatively small and clearly defined populations of healthy adults which are not representative of the broader residential populations typically associated with epidemiological studies, including older individuals and individuals with existing health conditions which place them at greater risk for O3-related effects.

However, the young adults studied in most of the human studies are the most sensitive to the first mild effects and studies of individuals with greater risk have not identified any greater responses. As noted above, Rom et al.[106] have pointed out that human clinical studies with exposures well above the current standard have been remarkably safe, even for members of sensitive subgroups. Rom et al. also point out that EPA has been sued over the fact that the human clinical studies expose subjects to dangerous and life-threatening levels of ozone and not informing the subjects of the possibility of death in the informed consent process. The Human Research Committees that oversee such studies would not have allowed them to occur if they thought they caused long-term harm or even mortality.

---

[105] PA, supra note 2, at Figure 4-13.
[106] Rom, supra note x.

The PA needs to acknowledge and discuss the dramatic disconnect between the mortality analyses and the human clinical risk assessment. In the clinical risk assessment, FEV1 decrements for the older individuals are so rare they are not even presented. Yet, the mortality analyses predict a substantial portion of respiratory mortality occurs at ambient concentrations for which the personal exposures of the population are even below the natural background. Currently, the draft PA ignores the stark difference and makes arguments about a continuum of effects from ozone. However, this line of reasoning totally ignores the issue of dose plausibility.

2. Observational Studies and Their Interpretation

In contrast to the human clinical studies that, if replicated, can establish cause and effect, the observational (or epidemiological) studies are more difficult to interpret. AIR reviewed the observational studies with respect to two considerations. The first is what the observational studies of the health endpoints identified in clinical studies - pulmonary function decrements, symptoms, and inflammation – can tell us about the effects of ozone on public health. The second consideration is whether the hospital admissions and mortality studies can inform us about the risk to public health.

a. Observational Studies of Clinically-identified Endpoints

(1). The Lung Function Data Are Less Consistent Than Claimed in the PA

Although there are many small positive associations of ozone with changes in lung function in the observational literature, the data are less consistent than indicated in the PA. The PA claims that observational studies have consistently linked increases in ozone to lung function decrements. However, the ISA acknowledges that the recent data is mixed, noting:

> Recent epidemiologic studies focused more on children with asthma rather than groups with increased outdoor exposures or other healthy populations. Whereas recent studies contributed less consistent evidence, the cumulative body of evidence indicates decreases in lung function in association with increases in ambient O_3 concentration in children with asthma. Collectively, studies in adults with asthma and individuals without asthma found both O_3-associated decreases and increases in lung function.[107]

Thus, the ISA indicates that a number of older studies comprise a majority of the supporting evidence from epidemiology regarding lung function test effects, whereas recent studies, provide less compelling evidence. In addition, the small changes in lung function that have been reported, to the extent they may be caused by ozone, are not medically significant given the transient, reversible nature of ozone-mediated lung function changes. The small changes and mixed results for asthmatic children are illustrated in Figure 6-7 of the ISA, where there are few statistically significant changes in FEV1.

---

[107] ISA, supra note 79, at p. 6-28.

A particular important study was carried out by the Health Effects Institute in the Los Angles Basin, the area of the country with the highest ambient ozone concentrations. Avol et al. (1998) concluded that the relationships between ozone and pulmonary function were erratic and difficult to reconcile with existing knowledge about the acute respiratory effects of air pollution.

(2). The Data on Inflammatory Markers and Respiratory Symptoms Is Inconsistent

The observational studies of ozone association with the presence of inflammatory markers or respiratory symptoms suffer from limitations due to the presence of other pollutants and multiple comparisons. The ISA also notes that the clinical relevance of most biomarker changes is not clear. The PA notes several additional reasons why there may be inconsistencies in the data. On balance, there was some evidence of associations of ozone with exhaled NO in Figure 6-11 of the ISA, but little consistency for other biomarkers. In addition, a number of these studies were conducted in Los Angeles and Mexico City where the subjects are exposed to high concentrations of both ozone and many other pollutants and report positive associations with various pollutants. Regarding inflammatory markers, the ISA indicates "The limited available evidence in children and adults with increased outdoor exposures and older adults was inconclusive."[108]

A particularly important study is described in the ISA as a well-designed panel study, Ferdinands et al. (2008). In this study, 16 adolescent long-distance runners in Atlanta, GA, were followed before and after exercise for 10 days in August 2004. Effect estimates for lags 0, 1, and 2 indicated O_3-associated <u>decreases</u> in airway inflammation. This study is important because the subjects, setting, and exercise level are just where one would expect to see ozone-induced inflammatory increases based on the clinical studies. Another study by Chimenti et al. (2009) measured some biological changes in adult male runners before and after races. However, the authors concluded that since no relationship was observed between neutrophil counts and inflammatory mediators 20 h after races, airways inflammation at this time point appears blunted in healthy runners and little affected by exposure to mild seasonal changes and airborne pollutants. Thus, under the situation with the greatest likelihood of inflammatory changes caused by ozone, there is little evidence of effects.

The lack of consistent increases in subclinical inflammatory markers is important information for the integrative synthesis. The lack of substantive effects in heavily exercising subjects suggests that there is even less likelihood of inflammatory changes due to ozone in the rest of the population as is goes about its daily activities. The findings in Adamkiewicz et al. (2004) of no inflammatory changes associated with ozone in elderly subjects including those with asthma and COPD confirm this view.

The evidence for respiratory symptoms associated with ozone in observational studies is mixed and inconsistent. For asthmatic children, the data appears somewhat consistent, but when one recognizes that similar data have been used by EPA to claim consistent

[108] Ibid., at p. 6-94.

effects on asthma from other pollutants, the reliance on single-pollutant studies is problematic. There are three multi-city studies that come to different conclusions with regard to individual pollutants. Therefore, the characterization of ozone having consistent effects on asthmatics cannot be supported. For children without asthma, the ISA acknowledges that the data are inconsistent, noting:

> Short-term increases in ambient O3 concentration were not consistently associated with increases in respiratory symptoms in groups comprising children with and without asthma.[109]

Although there are some positive associations with all the clinically-identified effects of ozone, there are also negative associations and null findings in the literature. The PA refers only to the positive single-pollutant associations, thereby giving a false impression of the overall data. The lack of consistent evidence implicating ozone as being associated with inflammation or respiratory symptoms in observational studies is an important finding that needs to be considered as the PA evaluates the biological plausibility of the more serious potential effects such as hospital admissions and mortality.

b. Observational Studies of Mortality and Hospital Admissions

The PA recognizes a number of major sources of uncertainty in evaluating the observational data. For example, the PA mentions heterogeneity in the ozone-mortality relationship across cities (or regions), methodological issues, and the question as to whether or not exposure errors, misclassification of exposure, or potential impacts of other co-pollutants may be obscuring potential population thresholds.[110] As AIR comments on the HREA demonstrate, all these issues are major impediments to assuming that ozone is causing mortality at concentrations below the current standard. In addition, despite decades of study of controlled exposures to ozone, there is a lack of experimental evidence that concentrations below the current standard can cause mortality. Given the full pattern of associations, not just EPA's preferred associations, the hypothesis that air pollution has no effect on mortality is entirely plausible. Therefore, the PA should not give any weight to the results of the mortality and morbidity risk assessment.

B. Prior CASAC Advice on the Level of the Standard is Not Particularly Relevant

The PA discusses CASAC advice in the 2008 review, as well as subsequent advice. Historically, CASAC provides the Administrator some mix of its collective and individual preferences for the range of alternative standards to consider in a NAAQS review.

The basic understanding of the effects of ozone from human clinical studies has not changed substantively since the 1997 review was completed. In that review, the Panel

[109] Ibid., at p. 6-121.
[110] PA, supra note 2, at p. 4-58 and 4-59.

acknowledged that there may be no threshold for ozone effects and, after evaluating the clinical and observational risk results, no CASAC panel member supported a standard level set lower than 80 ppb. The Administrator set an 80 ppb standard, citing uncertainties in the health evidence for exposure to concentrations below 80 ppb, the advice of CASAC, and the fact that a standard set at a level of 70 ppb would be closer to peak background concentrations that infrequently occur in some areas due to nonanthropogenic sources of O_3 precursors.

In the next review, one important consideration that changed substantively was the consideration of the background of ozone uncontrollable through control of U. S. emissions. EPA switched from using ozone measurements in remote locations to estimate peak background to using photochemical modeling to estimate mean "policy relevant" background. In addition, EPA claimed that peak background levels do not coincide with man-made ozone peaks. While the ISA and PA now confirm that this view of background was incorrect, CASAC, based on these false assumptions, indicated that the Panel preferred an 8-hour standard between 60 and 70 ppb.

In the current review, the estimates of risk are similar to those in both the 1997 and 2008 reviews but there are now improved estimates of mean USB and EPA acknowledges that peak background can reach between 60 and 75 ppb. Therefore, a 70 ppb standard could be confounded by peak background. For these reasons, prior CASAC preferences for the level of the standard should not be given any substantial weight.

C. Adequacy of the Current Standard

This section of the PA addresses the important question of whether it is appropriate or necessary for the Administrator to revise the existing primary ozone standard. Based on AIR's review of the ISA, HREA and PA, the current standard is highly protective of public health. The arguments mustered in the draft PA strain to make the case that the current standard is inadequate. As noted above the evidence of effects based on the clinical studies is similar between the 1997, 2008, and the current review. Also in each case, selected observational associations were used to estimate hospital admissions and now mortality due to ozone. However, it has become clear that the observational data are very noisy, with substantial stochastic variation. In multi-city studies, the range of single-city associations from strongly positive to strongly negative is biologically impossible. It is also clear that model selection uncertainty, publication bias, and potential confounding limit the interpretation of the hospital admissions and mortality associations as real health effects. The assumption in the HREA that ozone causes such effects down to zero concentrations is in contradiction of the known effects of ozone in human clinical studies and cannot be scientifically supported. Therefore, the results of the observational risk assessment are too uncertain to use in establishing the primary standard.

The PA discusses the new evidence for effects below 80 ppb, indicating:[111]

[111] PA, supra note 2, at p. 3-120.

> Compared to the evidence available in the last review, these studies have strengthened support for the occurrence of abnormal and potentially adverse respiratory effects following short-term exposures to O_3 concentrations below 80 ppb. It is reasonable to judge exposures to such O_3 concentrations to be potentially important from a public health perspective given the following:
>
> 1. The respiratory effects reported following exposures to O_3 concentrations of 60 and 70 ppb, while at moderate exertion, can reasonably be judged adverse based on ATS criteria and past advice from CASAC.
>
> 2. The controlled human exposure studies reporting these respiratory effects were conducted in healthy adults, while at-risk groups (e.g., asthmatics) could experience larger and/or more serious effects.
>
> 3. These respiratory effects are coherent with the serious health outcomes that have been reported in epidemiologic studies (e.g., respiratory-related hospital admissions, emergency department visits, and mortality).

Each of these three points will be discussed in turn.

Regarding the first point, the new data at 60 and 70 ppb only refine the dose-response curve at these concentrations. In previous reviews the Agency and CASAC assumed there were responses in this range. The mild responses at these concentrations with extended exercise levels (that are at the extremes of activity levels in the population) are not adverse using prior guidance or ATS criteria. That is why the PA refers to the effects as "abnormal and potentially adverse." Based on the HREA clinical risk assessment, the mild responses occur to only a small portion of the population the order of once per year. In prior reviews, isolated FEV1 decrements that are transient and reversible have not been considered adverse. The overall risk in terms of fraction or percent of person-days is extremely small, the order of a fraction of one percent.

The second point is entirely speculative. The available data indicate that older individuals and COPD patients have lesser responses and asthmatics have similar responses to the normal population.

Regarding the third point, the small respiratory effects predicted by the HREA are not coherent with the serious health outcomes predicted by EPA's favored epidemiological associations. Coherence as understood in the Bradford Hill criteria for assessing cause and effect means that an association should be in line with substantive knowledge or should not conflict with substantive knowledge. Another way of stating this is if laboratory experiments in which variables are controlled and external everyday evidence are in alignment, then it is said that there is coherence. Clearly, the laboratory experiments of ozone exposures to humans (or animals) do not demonstrate dire effects at low exposures.

Since the PA arguments for inadequacy of the current standard fail, the Administrator

should be informed that the current standard is highly protective and consideration of confirming that standard is entirely appropriate, especially considering that the current standard is very close to the uncontrollable background of ozone.

D. Comments on Chapter 4 – Consideration of Alternative Primary Standards

The fundamental question is what is the range of potential alternative standards that are supported by the currently available scientific evidence and exposure/risk information? AIR supports the view that the current indicator, ozone, the current averaging time, 8 hours, and the current statistical form are still appropriate. This leaves the question of the level or concentration of a primary standard that will protect the public health.

In discussing the human clinical studies, the PA overstates the public health significance of effects reported in 8-hour exposures below 80 ppb with exercise. The PA also overstates the number of such exposures, as documented above in comments on the HREA. Finally, the PA does not present the percent of person-days with various FEV1 decrements. As a result of the overstatements and omissions, the PA overstates the public health implications of either the current standard or alternative standards.

In discussing the observational studies, the PA focuses on finding studies with positive ozone associations in locations with the lowest ozone concentrations. Given the extensive noise or stochastic variation in the data, this exercise will locate the outliers in the full range of observational data and, as such, is not a reliable way to establish air quality standards. In addition, the HREA estimated mortality and morbidity effects using EPA's favored associations and assumptions. As documented above, this leads to the counterintuitive result that even complete elimination of U. S. man-made ozone precursors will leave a residual mortality burden due to ozone. The PA acknowledges many of the uncertainties in the interpretation of the observational studies. Therefore, the observational studies are not useful in choosing the level of the standard.

As part of the HREA, the Agency carried out photochemical modeling to determine the distribution of ozone concentrations upon attaining various alternative standards. One of the by-products of this exercise is that is provides estimates of the degree of additional emission control needed to attain the current standard and alternative standards. The exercise demonstrates that attaining even the current standard will require massive additional emission control in some locations. Attaining the alternatives evaluated in the PA will require an even greater degree of emission control.

The policy makers and readers of the PA need to know that the U. S. has been controlling ozone precursor emissions for over 50 years, with dramatic reductions in peak ozone levels in and downwind of cities, particularly major cities. However, due to the presence of a substantial background of uncontrollable ozone, there has been little change in the more remote areas of the U. S. After an enormous long-term national effort to reduce ozone, the EPA modeling projects that a further massive reduction will be needed to attain either the current standard or any of the alternative standards. In contrast, the risk

reduction in terms of public health, even using the HREA assumptions that overstate the risk, is minimal.

Therefore, there is no "bright line" that distinguishes any of the proposed standards as being significantly more protective of public health. This is the same judgment that the CASAC panel made during the 1997 review when considering similar probabilistic exposure estimates of FEV1 decrements and estimates of asthma hospital admissions based on observational studies.[112]

In addition, alternatives lower than the current standard are close to and may be exceeded by background ozone. Therefore, the range of alternatives for the Administrator to consider should include the current standard.

In the following sections, AIR concerns in each of these areas are documented.

1. The PA overstates the public health significance of effects reported in 8-hour exposures below 80 ppb with exercise

The PA strains to make the case that the effects at 60 ppb are adverse. For example, it claims that the small decrease in mean FEV1 meets the ATS criteria for an adverse response given that a downward shift in the distribution of FEV1 would result in diminished reserve function, and therefore would increase risk from further environmental insult.

In contrast, Goodman et al. (2013)[113] evaluated the effects below 80 ppb and concluded:

> In summary, the small decrements in pulmonary function, as represented by slightly decreased mean FEV1 values with no or slight concomitant changes in FVC, observed at relatively low ozone concentrations, are of low severity because they do not interfere with normal activity and do not result in permanent respiratory injury or progressive respiratory dysfunction. In addition, because the decrements in FEV1 and FVC are reversible, transient and represent a reflexive nervous response, these small changes represent a lesser degree of adversity than irreversible and sustained changes in cellular composition or in lung function.

The small transient and reversible mean FEV1 decrements at 60 ppb caused by a reflexive response that last only a few hours are not the kind of permanent shift that the ATS expresses concern over.

The PA also argues that a 10% decrement in FEV1 is accepted by ATS as an abnormal response, and based on advice received from CASAC in previous reviews, such

---

[112] G. T, Wolff, November 30, 1995 letter to Administrator Browner, EPA-SAB-CASAC-LTR-96-002, CASAC Closure on the Primary Standard Portion of the Staff Paper for Ozone.

[113] J. E. Goodman, R. L. Prueitt, J. Chandalia, S. N. Sax. Evaluation of adverse human lung function effects in controlled ozone exposure studies. Journal of Applied Toxicology on-line July 9, 2013.

decrements could be adverse in people with lung disease. However, the ATS guidelines for adverse effects do not specify a 10 % cutpoint and caution against considering FEV1 decrements by themselves as adverse. In previous reviews, isolated FEV1 decrements even with mild symptoms were not considered a concern.

At the 70 ppb level, the PA argues that there is evidence of respiratory symptoms combined with lung function decrements which would be an "adverse" response based on ATS criteria. However, the group mean FEV1 responses at 70 ppb are small and the symptom responses are also mild; both return to normal within a few hours or less after exposure. As noted above, in previous reviews, isolated FEV1 decrements even with mild symptoms were not considered a concern.

The PA also raises the issue of ozone-induced pulmonary inflammation at 60 ppb (Kim et al., 2011), noting inflammation is evidence that injury has occurred and raising the concern that repeated events of acute inflammation can have several potentially adverse outcomes. However, the extent of inflammation at 60 ppb is small as shown in Figure 11 below which is taken from Kim et al., 2011. In addition, Pino et al.[114] concluded that "…neutrophils do not play a detectable role in contributing to the early epithelial damage in the lung caused by an acute exposure to ozone" based on experiments with neutrophil-depleted rats exposed to 1,000 ppb ozone.

The immune system responses noted in the PA as the first indications of "inflammation" are physiological processes that occur in all living organisms under the stimuli of daily life. The first reported changes are small and reversible and well within the range of physiological variability. They fall into the category of biochemical markers that the American Thoracic Society indicates do not necessarily imply adversity. EPA has noted that the initiation of inflammation is an important component of the defense process; however, its persistence and/or its repeated occurrence can result in adverse health effects.

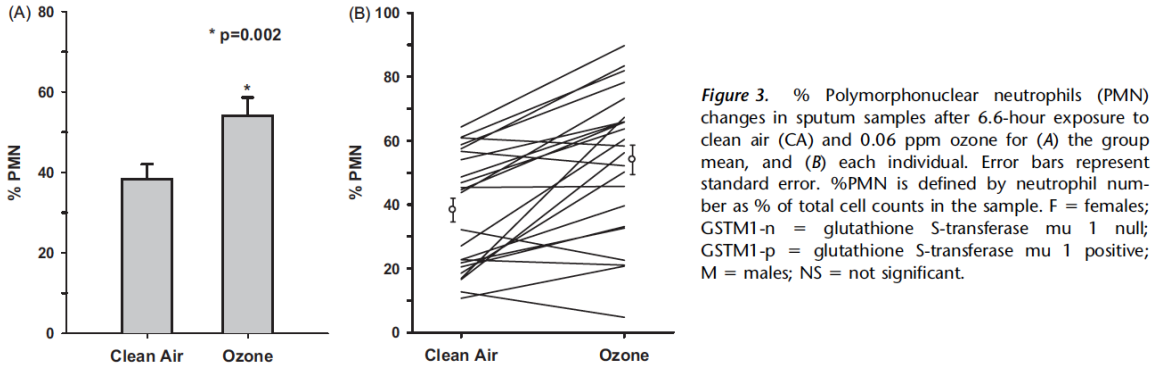

Figure 3. % Polymorphonuclear neutrophils (PMN) changes in sputum samples after 6.6-hour exposure to clean air (CA) and 0.06 ppm ozone for (*A*) the group mean, and (*B*) each individual. Error bars represent standard error. %PMN is defined by neutrophil number as % of total cell counts in the sample. F = females; GSTM1-n = glutathione S-transferase mu 1 null; GSTM1-p = glutathione S-transferase mu 1 positive; M = males; NS = not significant.

Figure 11. Neutrophil changes at 60 ppb ozone from Kim et al. 2011.

A recent human clinical study sponsored by the California Air Resources Board, that was

[114] Pino et al. Acute ozone-induced lung injury in neutrophil-depleted rats. Toxicology and Applied Pharmacology Volume 114, Issue 2, June 1992, Pages 268–276.

designed to test for systemic effects of ozone on inflammation in order to evaluate potential cardiovascular effects from ozone, found instead that the exercise intensity used in the human clinical studies produced an acute systemic inflammation that was of the same order of magnitude as the acute lung inflammation reported by Kim et al. 2011.

Balmes et al. report:

> In this research project, we first looked at the effects of intermittent moderate-intensity exercise as frequently used in human inhalational exposure studies on a variety of endpoints of relevance to cardiovascular responses to O3 exposure. We found that exercise induced a systemic pro-inflammatory response characterized by an immediate post-exercise increase in peripheral blood monocyte, neutrophil, and lymphocyte counts; an immediate increase in serum IL-6 and MCP-1 concentrations; and a delayed increase in serum CRP at 24 h post-exercise.[115]

When 100 ppb and 200 ppb ozone were added to the 4-hour exercise-only exposures, there was little evidence of an increased systemic inflammation and little evidence of any other cardiovascular-related response. Thus, the stress of vigorous exercise, by itself, produces an acute pro-inflammatory response. Therefore, the initial ozone pro-inflammatory responses below the current standard should not be considered a threat to public health.

The PA also overstates the number of exposures of concern and of potential FEV1 decrements, as documented above in comments on the HREA. In order to put these responses into perspective, the PA should also provide the percent of person-days with various FEV1 decrements. Using the counts in Appendix 6B and the population and ozone season data in Table 5-1 of the HREA, AIR prepared the following Figure 12 which is analogous to Figure 4-5 of the PA.

[115] J. R. Balmes, M. Arjomandi, H. Wong, A. Donde, and K. Power, Effects of Ozone Exposure on Cardiovascular Responses in Healthy and Susceptible Humans, California Air Resources Board Contract Number 04-322, October 2011.

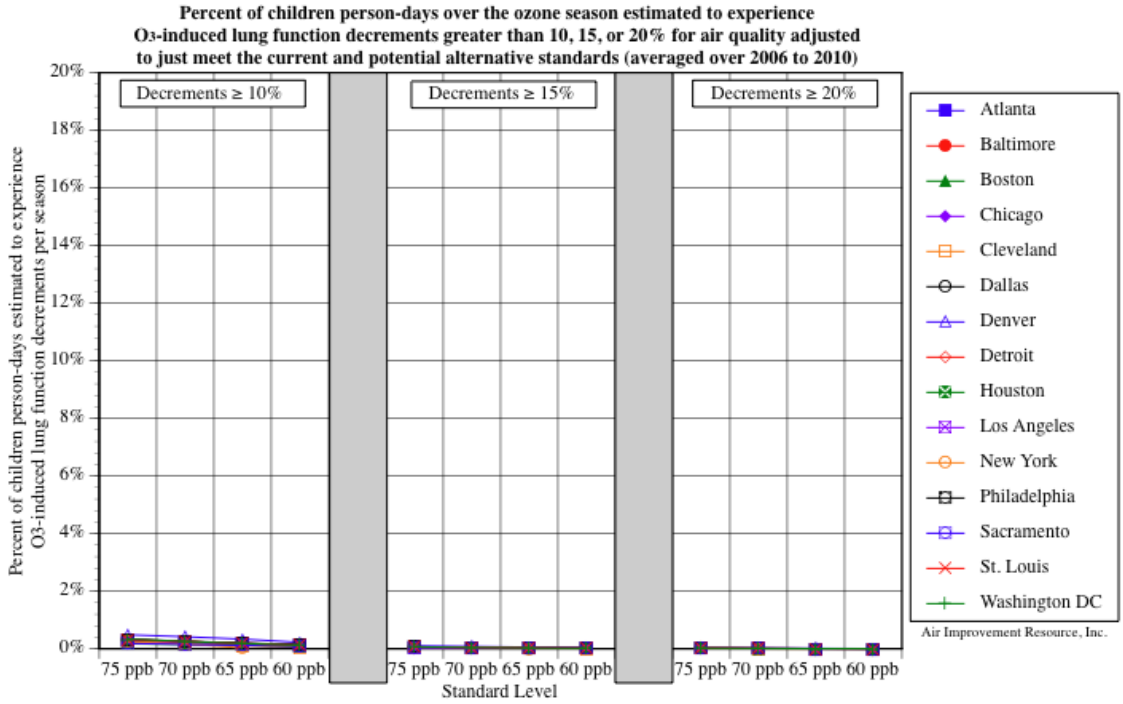

Figure 12.

Since the percent of person-days of occurrences is vanishingly small, the same data is presented in Figure 13 with the x-axis expanded. Clearly any of the alternative standards is extremely protective, with only a portion of a percent of the total exposures resulting in an exposure of any potential concern.

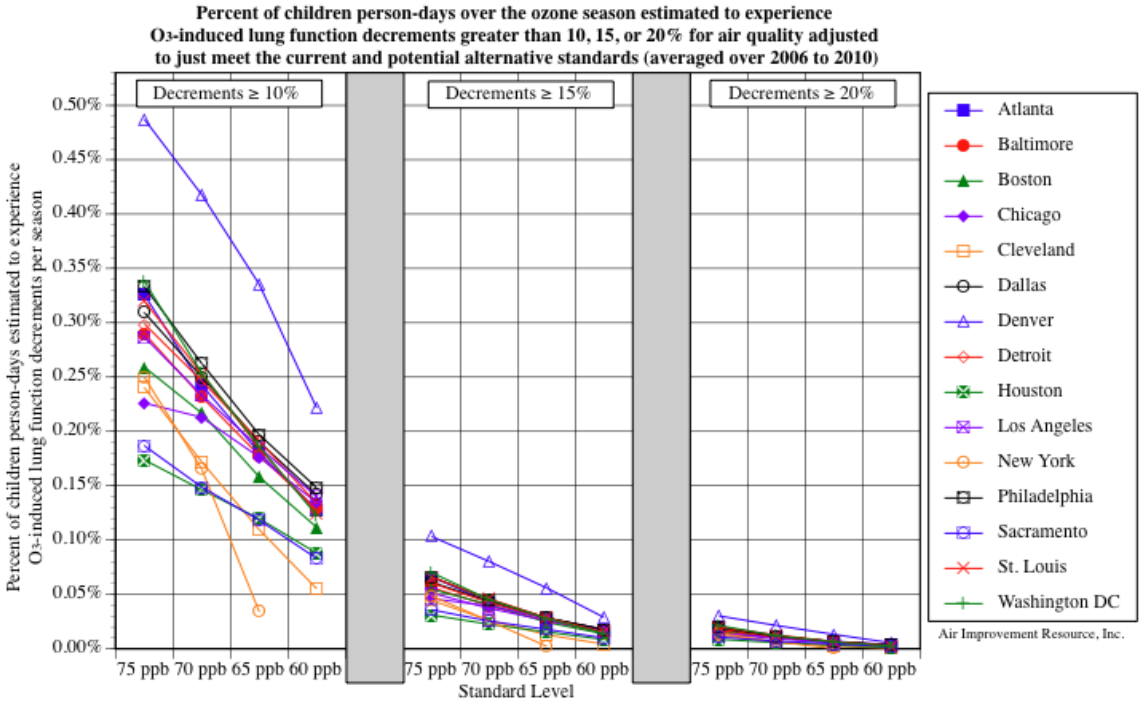

Figure 13.

As discussed above and documented in the PA, the MSS FEV1 responses, which are displayed in Figures 12 and 13, were about twice as numerous as the E-R FEV1 responses. Due to the noise in the underlying FEV1 data, the MSS model predicts FEV1 decrements a small fraction of the time even at low ozone exposures and low levels of exercise. Since both MSS and E-R FEV1 decrements are based on APEX exposure and ventilation rate estimates, both are also biased high because of the three factors discussed in Section II.A. Therefore, the estimates in Figures 12 and 13 should be considered an upper limit of the potential risks of FEV1 decrements. Even at the upper limit, the risk is minimal.

As a result of the overstatements and omissions AIR has identified, the PA overstates the public health implications of either the current standard or alternative standards.

2. The observational studies are not useful in choosing the level of the standard

In discussing the observational studies, the PA focuses on finding studies with positive ozone associations in locations with the lowest ozone concentrations. Given the extensive noise or stochastic variation in the data, this exercise will locate the outliers in the full range of observational data and, as such, is not a reliable way to establish air quality standards. In addition, the HREA estimated mortality and morbidity effects using EPA's favored associations and assumptions. As documented above, this leads to the counterintuitive result that even complete elimination of U. S. man-made ozone precursors will leave a residual mortality burden due to ozone. The PA acknowledges many of the uncertainties in the interpretation of the observational studies. Therefore, the observational studies are not useful in choosing the level of the standard.

3. Fifty years of emission controls have reduced peak ozone in and downwind of cities but not in remote areas

As part of the HREA, the Agency carried out photochemical modeling to determine the distribution of ozone concentrations upon attaining various alternative standards. One of the by-products of this exercise is that is provides estimates of the degree of additional emission control needed to attain the current standard and alternative standards. The exercise demonstrates that attaining even the current standard will require massive additional emission control in some locations. Attaining the alternatives evaluated in the PA will require an even greater degree of emission control.

The policy makers and readers of the PA need to know that the U. S. has been controlling ozone precursor emissions for over 50 years, with dramatic reductions in peak ozone levels in and downwind of cities, particularly major cities. The PA points out that the first Federal ozone-related standard was set in 1971. However, the nation's emission control program started before the 1971 air quality standard. Concern over ozone and photochemical pollution started in Los Angeles in the 1950s and led to the first auto emission controls (positive crankcase ventilation) on new cars in 1961 in California and 1963 nationwide. Exhaust emission standards began in 1966 in California and in 1968 nationally. Unleaded gasoline was made available in 1971 at the request of the auto

industry and, in 1975, the catalytic convertor became part of the emission control system. There have been many rounds of ever-tighter vehicle emission standards for ozone precursors in the intervening years. There have also been regulations for many other VOC and NOx sources as well. Thus, there have been 50 years of national as well as regional, and local control programs aimed at reducing ambient ozone in the U. S.

The reason this history is relevant to the current review is that it represents a 50-year long real-world experiment that can tell us something about the portion of the ozone that has changed in response to the emission reductions. This in turn can provide insight into the portion of the ozone distribution that is amenable to further control. There are two important findings from the various trend studies that have been carried out by EPA and others. The first is that 50 years of controls have primarily reduced peak ozone in and downwind of cities. The second is that, due to the presence of a substantial background of uncontrollable ozone, there has been little change in the more remote areas of the U. S.

For example, Figures 14 and 15 show the trend in peak 1-hour and 8-hour ozone concentrations at monitoring sites across the country from 1975 to 2010. Figures 16 and 17 show the trend in June through August average daily maximum 1-hour and 8-hour ozone concentrations from 1975 to 2010. These figures show the progress that the nation's ozone control program has made. There has been dramatic progress in reducing the 1-hour and 8-hour daily maxima at the highest ozone that which are dominated by man-made ozone. However, the more limited progress at lower ozone sites is indicative of the presence of a substantial background of uncontrollable ozone as documented in Section I of these comments.

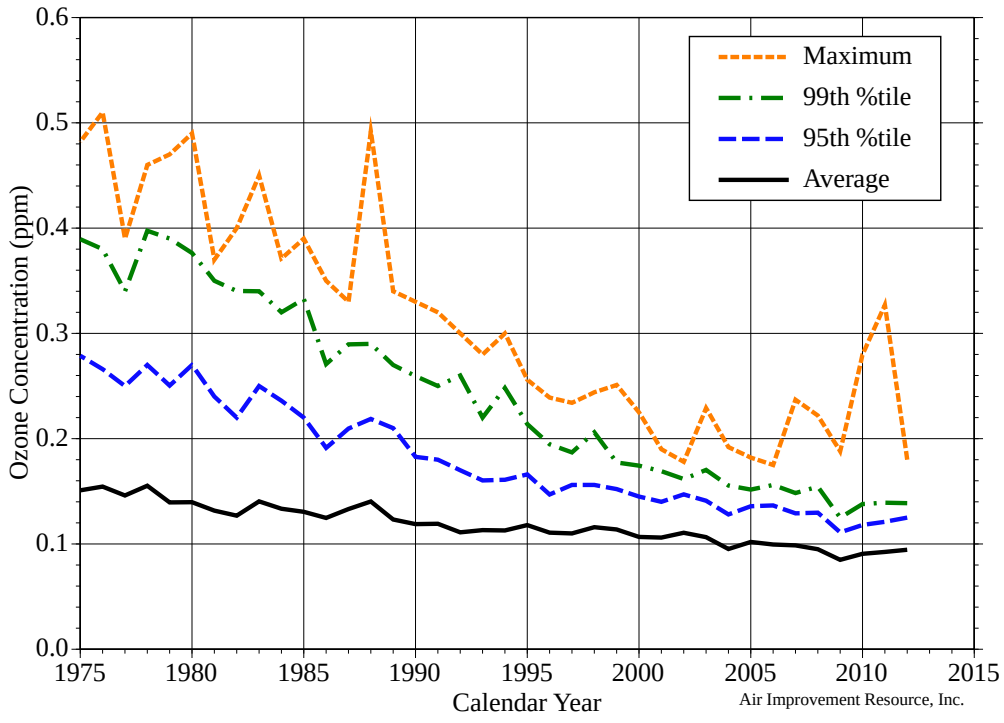

Figure 14: Highest annual 1-hour ozone concentrations for all US monitoring locations from 1975 to 2012.

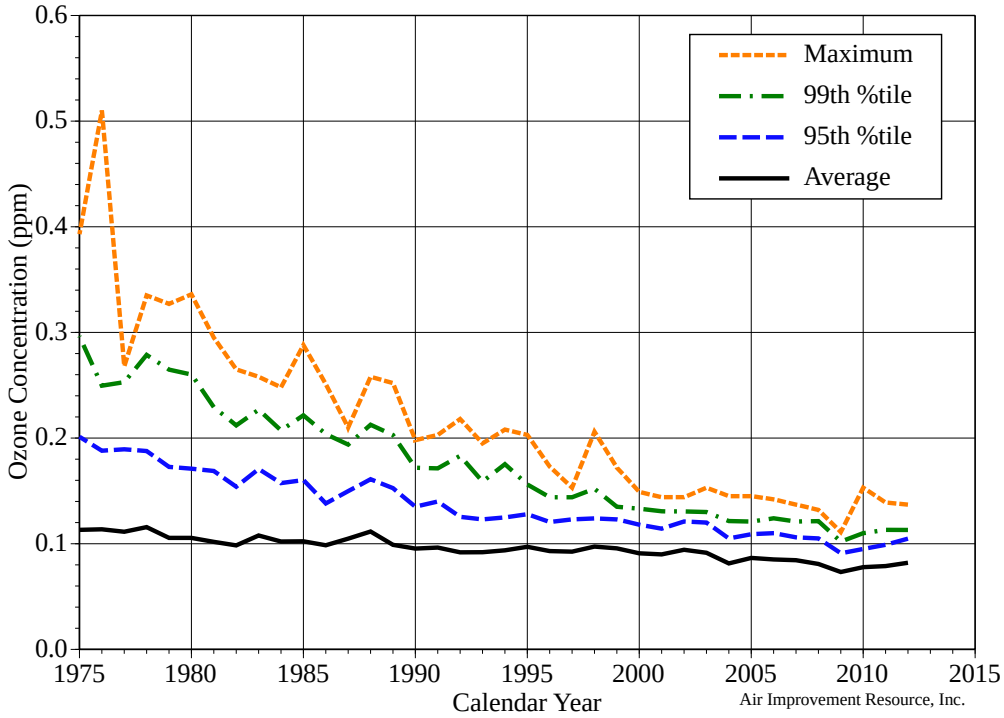

Figure 15: Highest annual 8-hour ozone concentrations for all US monitoring locations from 1975 to 2012.

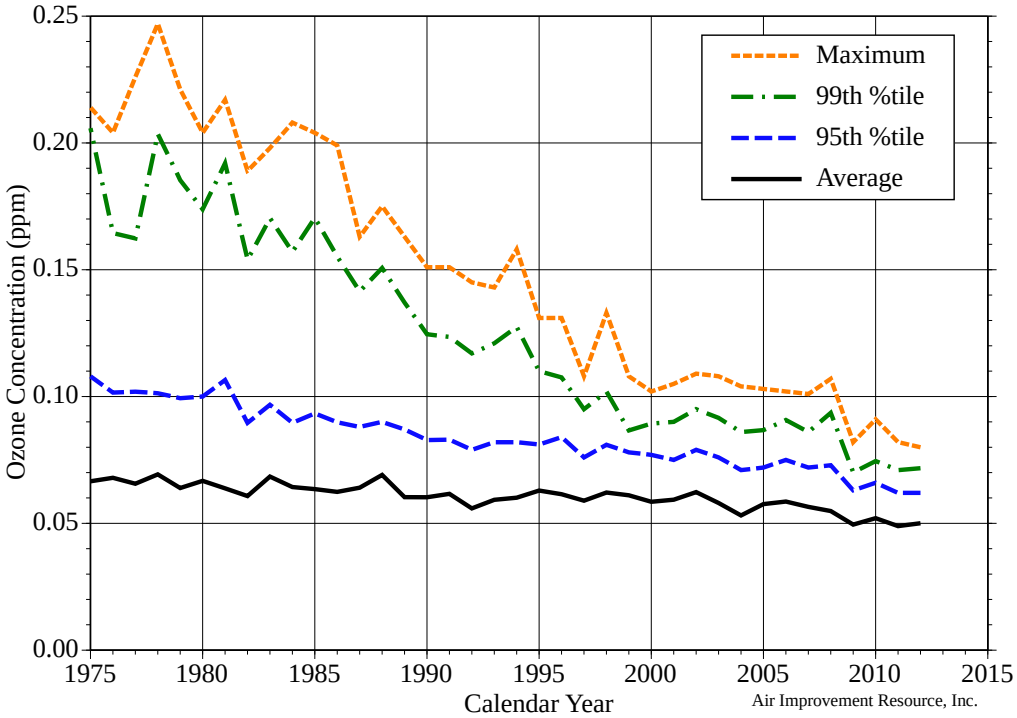

Figure 16: June - August average daily peak 1-hour ozone concentrations for all US monitoring locations from 1975 to 2012.

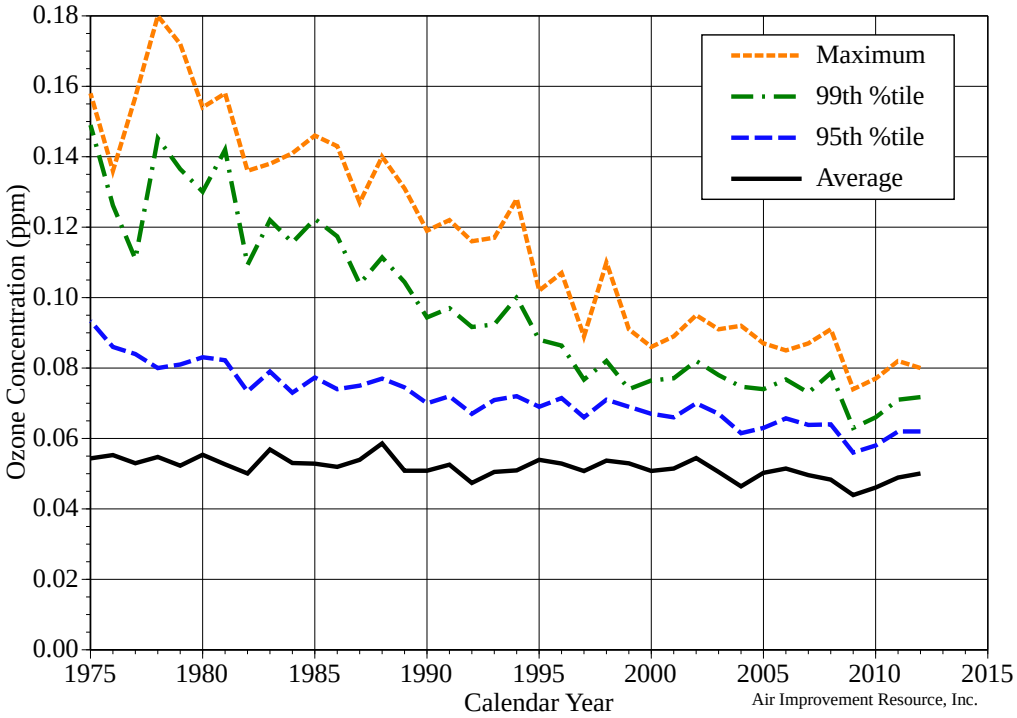

Figure 17: June - August average daily peak 8-hour ozone concentrations for all US monitoring locations from 1975 to 2012.

An EPA ozone trends plot of the distribution of 8-hour design values, shown in Figure 18, from 1980 to 2012 show a similar behavior. There is a substantial portion of the ozone which has not been amenable to control.

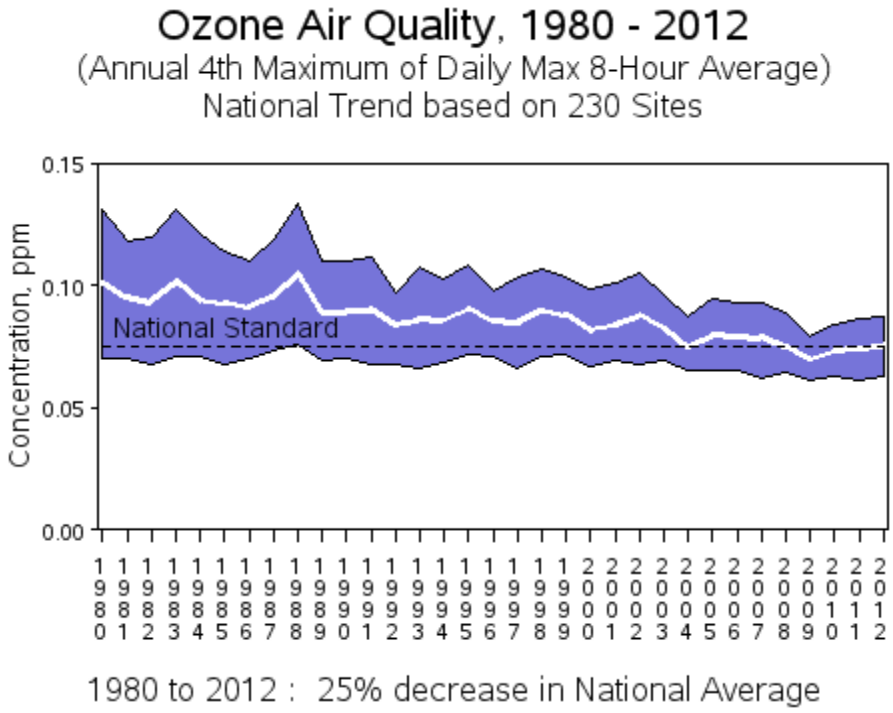

Figure 18.

The ozone trends at rural and remote sites in the U. S. have recently been evaluated by Cooper et al.[116] In their introduction, Cooper et al. refer to several studies that show, as demonstrated above, that U.S. emission controls are reducing the frequency and magnitude of extreme ozone episodes. However, Cooper et al. also caution that while emission reductions appear to be reducing the frequency of high ozone events, several studies have shown that mixing ratios of the lower ozone percentiles, such as the 5th, 10th or 20th percentiles are increasing across the country.

Cooper et al. analyzed long-term (1990–2010) rural ozone trends using all available data in the western (12 sites) and eastern (41 sites) US. To minimize local effects and to ensure that the ozone measurements are representative of the well-mixed daytime atmospheric boundary layer, they used only hourly average ozone data from 11 am local time to 5 pm local time. Rather than focus on average ozone or air quality standard violations, they considered the full range of ozone values, evaluating trends for the 5th, 50th and 95th percentiles.

Based on EPA inventories, domestic ozone precursor emissions for VOC, CO, and NOx each decreased by roughly half from 1990 to 2010. Cooper et al. report that 83%, 66% and 20% of summertime eastern U.S. rural sites experienced statistically significant ozone decreases in the 95th, 50th and 5th percentiles, respectively. During spring 43% of the eastern sites had statistically significant ozone decreases for the 95th percentile with no sites showing a significant increase. At the 50th percentile there was little overall change in the eastern U.S.

In contrast, only 17% (2 sites) and 8% (1 site) of summertime western U.S. sites had statistically significant ozone decreases in the 95th and 50th percentiles, respectively. During spring no western site had a significant decrease, while 50% had a significant median increase. Cooper et al. discuss the dichotomy in U.S. rural ozone trends in terms of changing anthropogenic and biomass burning emissions and the probability that increasing baseline ozone flowing into the western U.S. is counteracting ozone reductions due to domestic emission reductions.

Cooper et al. also evaluated the trend in springtime free tropospheric ozone (between 3 and 8 km above sea level) above western North America using all available data from ozonesondes, research aircraft and commercial aircraft. They report that April–May free tropospheric ozone increased from 1995 to 2011 at the rate of 0.4 ppb per year. The median springtime free tropospheric values over the western North America are about 60 ppb, while the 5^{th} and 95^{th} percentile values are 30 and 90 ppb, respectively.

The lack of trend in western rural sites together with the presence of a large and slowly growing reservoir of ozone in the free troposphere suggests that it will be very difficult if not impossible to lower ozone in the rural areas of the intermountain west, where current

[116] Cooper, O. R., R.-S. Gao, D. Tarasick, T. Leblanc, and C. Sweeney (2012), Long-term ozone trends at rural ozone monitoring sites across the United States, 1990–2010, J. Geophys. Res., 117, D22307, doi:10.1029/2012JD018261.

design values are in the range of 60 to 75 ppb as shown in Figures 2-3 and 2-4 of the PA.

After an enormous long-term national effort to reduce ozone, the EPA modeling projects that a further massive reduction will be needed to attain either the current standard or any of the alternative standards in and downwind of major cities. In contrast, the risk reduction in terms of public health, even using the HREA assumptions that overstate the risk, is minimal.

4. There is no "bright line" that distinguishes any of the alternative standards as being significantly more protective of public health

With regard to the human clinical studies, when both metrics of response - percent of persons with one or more occurrences and percent of person-days of occurrence -are considered, there is no "bright line" that distinguishes any of the proposed standards as being significantly more protective of public health. As noted above, this is the same judgment that the CASAC panel made during the 1997 review when considering similar probabilistic exposure estimates of FEV1 decrements.

With regard to the observational studies of endpoints such as hospital admissions and mortality, even if one accepts the EPA favored associations and assumptions, the impact of alternative standards is quite small as a portion of the total incidences. In addition, the assumption of no threshold is not consistent with the large body of controlled exposure studies. Since the full pattern of associations for these endpoints is not consistent with causality, the observational studies are not useful in choosing the level of the standard.

Finally, alternatives lower than the current standard are close to and may be exceeded by background ozone, as acknowledged in the PA. There is precedent for considering background in ozone NAAQS decisions. The 80 ppb 1-hour photochemical oxidant standard was revised to a 1-hour ozone standard of 120 ppb in 1979, in part, because there was evidence that it was too close to background. In the 1997 review, an 8-hour standard of 70 ppb was viewed as being too close to peak background. Therefore, the range of alternatives for the Administrator to consider should include the current standard.

Appendix 1 -- APHENA O_3 Comments

The combined results of the large and comprehensive APHENA study are not consistent with ozone having a causal role in mortality or morbidity below the current standard.

In October, 2009, the Health Effects Institute (HEI) published the results of the *Air Pollution and Health: A European and North American Approach (APHENA)*[117] study. The APHENA project was designed to take advantage of the largest databases available. These had been developed by the three groups of investigators for earlier studies: 1) the *Air Pollution and Health: A European Approach* Phase 2 (APHEA2) study involving 32 cities; 2) the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), conducted in the 90 largest U.S. cities; and 3) multicity research on the health effects of air pollution in 12 Canadian cities. Each database included air pollution monitoring data for particulate matter and ozone, health outcome data in the form of daily mortality for all ages, for persons younger than 75 years, and for persons 75 years or older (from all nonaccidental causes [all cause]), cardiovascular disease, or respiratory disease) and daily hospital admissions for persons 65 years or older (for cardiovascular and respiratory disease). Other database variables used for APHENA included weather data and a number of socioeconomic and other variables known or suspected to influence mortality or hospital admissions.

In the original studies, each of the three groups used different modeling methodologies and entered different variables into their models. Although each group found positive and significant relationships between PM_{10}/O_3 and mortality and some morbidity endpoints, the magnitude of the relationships differed by geographic region. One goal of APHENA was to use common methodologies and variables and reanalyze their data sets. They intended to create a central repository for all three of the time-series databases and use a common quality assurance approach. In addition, they would conduct analyses on a combined, pooled dataset to study a variety of sensitivity issues including effect modification. They would then investigate the sensitivity of the estimates to a variety of smoothing methods and to the number of degrees of freedom. They also intended to explore reasons for the geographical heterogeneity of the effect estimates seen in their original studies. Another important goal of the program was to understand the extent of coherence between mortality and hospitalizations using data from cities in North America and Europe.

In the original analyses, all three groups used a two stage approach. In the first stage, risks were estimated for the individual cities, and in the second stage, evidence across the cities were combined. Each group used different methods to perform both stages in the original analyses. In APHENA, the investigators wanted to identify a preferred way to do both stages and apply common methodologies to the three data sets. For the first

[117] Katsouyanni K. and Samet, J. (2009). Air Pollution and Health: A European and North American Approach (APHENA), HEI Report 142, October, 2009.

stage, they identified two smoothing techniques, natural splines (NS) and penalized splines (PS), and decided to use a number of degrees of freedom choices. They chose to use 3, 8 and 12 degrees of freedom and also the number of degrees of freedom chosen by minimizing the partial autocorrelation function (PACF).

For the second stage analyses, the two approaches used in original NMMAPs and the European studies represented the two major approaches used at the time to pool estimates. NMMAPS used Bayesian hierarchical regressions models while the Europeans used metaregression models. However, they could not determine which was the best method, so they decided to use the models interchangeably.

Using the two smoothing techniques together with the four choices for the degrees of freedom and three choices of lags (0-1 day, 1 day and distributive lags which provided the cumulative effects of days 0 through 2) for each health outcome, the investigators ran a total of 24 different models for ozone. In addition, subsets of these choices were also used to examine the effects of controlling for PM_{10} and seasonal variations.

The results showed that the differences between the PS and the NS were very small in most cases and that the number of degrees of freedom tended to give similar results when greater than 6-8 degrees of freedom where used.

The overall modeling results for the mortality models and the morbidity models are summarized in Table 1 and 2, respectively. The denominator in the tables is the total number of different models that were run for each health effect outcome examined and the numerator is the number of models that resulted in a positive and statistically significant relationship between ozone and the health effect outcome. The way to interpret these tables is as follows. High ratios are suggestive of a robust and consistent relationship while low ratios are suggestive of no significant relationship. Intermediate values of the ratio suggest inconsistent and non-robust relationships that are dependent upon the model selected. Since there is no a priori way to determine the "correct" model, it is not possible to determine whether a small number (low ratio) significant and positive relationship represents real causal relationship or if they are false positives that can occur by chance or by confounding.

The all cause, all ages mortality results indicate a consistent relationship with ozone in Canada but somewhat less consistent relationships in Europe and the US. When the results for the two different age groups are examined, the interpretation of the results becomes even less clear. For $\geq$ 75 years of age, a consistent relationship still holds in Canada, but the European and US relationships become less consistent. When compared to the results for the < 75 years of age group, the results are implausible as they suggest that ozone is affecting the younger group more than the older group which goes against conventional wisdom. Controlling for PM makes the positive relationship for the older group disappear in all three locations, but the positive effect remains for the younger group except in the US where no relationship is evident. At all three locations a consistent summertime relationship is seen but vanishes in Europe and the US when PM is controlled. PM controlled model results were not presented for the Canadian data. In

any event, the results are not consistent with the existence of a causal relationship between ozone and all cause mortality.

The cardiovascular mortality/ozone modeling results are somewhat confusing. A clear positive relationship was found only in Canada and only for the ≥ 75 years of age group. Few significantly positive relationships were found for either age group for the other locations and no relationship was found in Canada for the younger age group. When PM is controlled for, few significant relationships remain. The summer only results suggest significant relationships in Europe and the US, but they vanish when PM is controlled. Taken altogether, these results do not support a causal relationship between ozone and cardiovascular mortality when the models are controlled for PM.

The cardiovascular hospital admissions/ozone results are also confusing. The annual results show a few significant model-dependent relationships in Canada and the US but none in Europe. When PM is controlled for, a few significant, model-dependent relationships remain in Canada, disappear in the US, but become consistently significant in Europe. The European results defy logic and were dismissed by the APHENA authors as a strong positive relationship was evident for respiratory hospital admissions and PM_{10}. The summer only results at all three locations show no significant relationships. Thus the weight of evidence from these results is consistent with the mortality results and does not suggest a causal relationship between ozone and cardiovascular hospital admissions.

In contrast to the cardiovascular mortality results, the respiratory mortality modeling results consistently show no relationship with one exception. None of the annual results at any location show any significant relationship between ozone and respiratory mortality. However for the summer, consistent significant results are found but only in Canada. Significant model-dependent results are seen in Europe and the US, but they disappear when controlled for PM. PM controlled results for Canada were not presented. Nevertheless, the weight of evidence of all the ozone/respiratory mortality model results does not support a causal relationship.

The respiratory hospital admissions show consistent significant relationships with ozone in Canada that disappears when PM is controlled. In the US and Europe, a few significant, model-dependent relationships are seen that persist when PM is controlled. However, during the summer when ozone is the highest and the strongest relationships would be expected, no significant relationships are found in either the US or in Europe. Consequently, the weight of evidence does not support a causal relationship between ozone and respiratory hospital admissions.

In summary, the APHENA results do not support EPA's claims of causal relationships between ozone and mortality or between ozone and hospital admissions.

Cause of Death	Canada	Europe	United States
All Cause – all ages	24/24	15/24	12/24
≥ 75 yrs	23/24	2/24	6/24
< 75 yrs	18/24	22/24	10/24
All Cause PM controlled – all ages	4/8	8/16	0/16
≥ 75 yrs	0/8	3/16	0/16
< 75 yrs	5/8	14/16	0/16
All Cause – summer only	9/9	18/18 (4/12)*	18/18(0/12)*
Cardiovascular – ≥ 75 yrs	24/24	3/24	2/24
< 75 yrs	0/24	8/24	2/24
Cardiovascular –PM controlled ≥ 75yrs	0/8	0/16	0/16
< 75 yrs	0/8	5/16	2/16
Cardiovascular – summer only	0/6	8/12(0/8)*	11/12(0/8)*
Respiratory – all ages	0/24	0/24	0/24
≥ 75 yrs	0/24	0/24	0/24
Respiratory – PM controlled – all ages	0/8	0/16	0/16
≥ 75 yrs	0/8	0/16	0/16
Respiratory – summer only	6/6	4/12(0/8)*	2/12(0/8)*

*Denotes the PM controlled ratio

Table A1. APHENA modeling results for mortality. The numerators represent the number of models that showed a positive and statistically significant relationship between O_3 and mortality while the denominator is the total number of models run.

Type of Admission	Canada	Europe	United States
Respiratory	18/24	8/24	7/23
Respiratory – PM controlled	0/8	7/16	5/16
Respiratory – summer only	3/3	0/4	0/4
Cardiovascular	5/24	0/24	3/24
Cardiovascular – PM controlled	3/8	16/16	0/16
Cardiovascular – summer only	0/4	0/4	0/4

Table A2. APHENA modeling results for hospital admission for patients 65 years and older. The numerators represent the number of models that showed a positive and statistically significant relationship between O_3 and admissions while the denominator is the total number of models run.