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"

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

Prepared for
The Alliance of Automobile Manufacturers

October 18, 2012

On September 11 and 13, we presented our preliminary comments to CASAC on the ozone Health REA and PA. Our written comments were submitted to EPA on Friday October 12. I have attached a copy of these comments and we would appreciate if you would distribute them to the CASAC Ozone Panel. In particular, we would like to call the Panel's attention to the following three issues:

1. Background ozone - The values EPA uses in the REA for U.S. background (USB) are unrealistically low compared to the values presented in the ISA (pages 13-15). In addition, the way EPA uses USB in the REA appears to conflict with the intent of CASAC (pages 21-22).

2. Interpretation of clinical studies - We would recommend that CASAC look at Tables 3 through 6. The APEX model provides two major outputs, the persons experiencing 1 or more exposures in a year to benchmark levels and the number of person-day occurrences at that level. We feel that the portion of person-days out of the total number of person-days is a more important metric for evaluating risk to public health. However, the first draft REA presents no data on person-days and includes lots of figures and tables for the number metric.

3. Interpretation of the epidemiology - The model selection uncertainty is very large, as documented in Figures 8 and 9. Figure 8 shows the unadjusted city-specific associations in Bell et al. (2004) and Zanobetti and Schwartz (2008) for the cities that the two studies have in common. There is little or no correspondence between the ozone/mortality associations in the two studies that EPA considers the best sources of data on this subject. Note also that there are many negative associations in the unadjusted data. Figure 9 shows a similar lack of correspondence in two separate NMMAPS analyses.

Thank you.

George T. Wolff, Ph.D.
Air Improvement Resource, Inc.

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"

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

Prepared for
The Alliance of Automobile Manufacturers

October 12, 2012

Executive Summary

Air Improvement Resource, Inc. (AIR) reviewed the first draft Health Risk and Exposure Assessment for Ozone (REA) and the first 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 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 assessment of risk which is then used to inform policy decisions that need to be made concerning the form, averaging time, and level of the NAAQS.

In the REA, EPA uses U.S. background (USB), which is defined as the ozone concentration that would exist in the absence of U.S. anthropogenic emissions. We agree that this is the appropriate background measure to use in the risk assessment. This is in contrast to the previous review which used North American background, which was defined as the concentration in the absence of U.S., Canadian and Mexican anthropogenic emissions. EPA claims to use the most

recent version of the photochemical grid model, GEOS-Chem, to estimate seasonal average values of USB for the 12 urban areas chosen for the risk assessment. However, the USB values presented in the REA are unrealistically low and the mean maximum daily 8-hour values for spring and summer are 10 - 15 ppb lower than the estimates from GEOS-Chem that are presented in the ISA and they are 15 - 20 ppb lower than those calculated from the preferred CAMx model. Using underestimated USB values, results in an overestimation of risks and an overestimation of the risk reduction that would be achieved when emissions are reduced.

In addition, they use a methodology in their assessment of risks that obscures the important contribution that background ozone makes to the risk estimates calculated by EPA. The contribution of risks occurring at concentrations of ozone at or below background becomes more important as lower NAAQS are considered. The methodology EPA uses to calculate risk assumes no threshold concentration for health effects and assumes that exposure to concentrations of ozone at or below background levels pose a real threat to human health. These assumptions also inflate the estimated health risks and the estimated health risk reductions when more stringent NAAQS are considered.

Consequently both the values of USB that EPA uses in the risk assessment and the methodology that EPA uses to calculate the risks are flawed, and both result in inflating the risk estimates.

Comments on the REA

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 – occur after exposures to 0.08 ppm 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 REA 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 in Denver and Los Angeles under current air quality is extremely small. For example for Denver in 2006, when the design value was 0.09 ppm, the portion of persons-days for children with FEV1 decrements >15 % was 0.00023 or 0.023 %. Thus using the REA methodology, current air quality is very protective of public health. Attainment of the current standard would reduce these already extremely small risks substantially. 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 second draft should correct the exposure model for the biases identified by AIR, include presentation of both persons and person-days results, and model the dose-response with a new threshold model that is available.

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 REA that dramatically overstate the

magnitude and certainty of ozone health risks. For example, the REA 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.

The REA presents a map of ozone mortality effects and a national mortality estimate of 18,000 premature deaths per year based on assuming a 0.5 % increase in mortality for a 10 ppb ozone increase in the previous week. AIR shows how the map is totally inconsistent with regional analyses, and how the national estimate is made up of city-specific estimates of risk that suggest ozone is associated with a biologically impossible range of effects from increasing mortality by up to 14 percent is some cities to decreasing mortality by up to 6 % in other cities.

By exploring the full range of spatial and temporal differences in association together with model selection uncertainty in the second draft REA, 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

Since the draft PA was written without any CASAC or public input on the REA or the final CASAC and public input on the Integrated Science Assessment, it represents solely EPA staff thinking. 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 2 of the PA, which summarizes the health evidence, 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 claims 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 second draft PA should address all these issues in the interpretation of the observational studies and their integration with the full range of ozone effects studies.

Chapter 3 just summarizes the first draft REA. The revisions that are necessary in the REA will have a major effect on the next draft of the 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.

Chapter 4 discusses the adequacy of the current standard and draws the preliminary conclusion that the current standard is inadequate to protect public health. AIR is concerned that without CASAC and public review of the REA, it is premature in the PA to draw any conclusions as to the adequacy of the current ozone standard.

The discussion of adequacy 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 for 8-hours at a rate that 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. Thus, these are rare occurrences at current ozone levels and will be even rarer occurrences when the current standard is attained.

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 studies have no threshold. Such an assumption is not consistent with either the general principles of toxicology or the specific findings of ozone toxicological studies. The 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) we now have 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.

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"

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

Prepared for
The Alliance of Automobile Manufacturers

October 12, 2012

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 first external review drafts of the Health Risk and Exposure Assessment for Ozone (REA)[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 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.

---

[1] U. S. Environmental Protection Agency, *Health Risk and Exposure Assessment for Ozone, First External Review Draft,* EPA-452/P-12-001, July 2012.
[2] 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.
[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.

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.

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 third draft ISA, this is the case and there is now substantial new modeling and other information that supports the Alliance view.

The human clinical studies of ozone are 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 now substantial information that the first physiological effects 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 REA includes probabilistic modeling of ozone exposures that attempts to answer this question. As documented in the following, the draft REA substantially overestimates the risk from the effects identified in the clinical studies.

The epidemiological or observational studies of the association of ozone with various health endpoints continue to be difficult to interpret. As more studies are published, the fundamental weaknesses of this body of information have become more apparent. Public comments from AIR and from several other scientists have detailed these concerns and inconsistencies.[5] However, the draft REA and PA continue to gloss over the issues that have been raised in public comments and fail to address the uncertainty and inconsistencies that are present in the epidemiologic data. As a result, the draft REA and PA overstate the consistency and weight of evidence for ozone effects from epidemiologic studies.

I. Background Ozone

---

[5] 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 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.

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 assessment of risk 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

1. Recent History

Since the release of the last Staff Paper (SP)[6] 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),[7,8,9] the subsequent Risk and Exposure Assessment (REA),[10] and the Policy Assessment (PA).[11] 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 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.[12]

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,[13] 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

---

[6] 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.
[7] 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, Mar. 2011.
[8] 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.
[9] 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.
[10] U. S. Environmental Protection Agency, supra note 1.
[11] U. S. Environmental Protection Agency, supra note 2.
[12] U. S. Environmental Protection Agency, supra note 7, at pp. 2-5.
[13] U. S. Environmental Protection Agency, supra note 8, at pp. 1-4.

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[14] 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[15], EPA had included three definitions of background ozone for consideration: NAB (as previously defined), USB and natural background. 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."

2. Recent Studies Raise Background Ozone Estimates

Since the last ozone review, there have been a number of field studies, data analyses and modeling studiesthat have shown that the PRB estimates that were used in the 2007 SP were unrealistically low and that USB, rather than PRB or NAB was the more appropriate measure of background ozone to use in the risk and policy assessments. These new studies have been highlighted in previous comments by AIR, Inc.[16,17,18] and others[19] and pertinent points will be briefly summarized below.

---

[14] 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.

[15] U. S. Environmental Protection Agency, supra note 9, at pp. 2-7.

[16] G.T. Wolff, 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.

[17] Heuss et al., supra note 14, at pp. 6-18.

[18] J. M. Heuss and G. T. Wolff, 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.

[19] N.D. Downey, D. Blewitt and D. Wood, Comments on EPA's Second Draft of the Integrated Science Assessment for Ozone and Related Photochemical Oxidants EPA/600/R-10/076B Released September 2011. Prepared for BP America, December 29, 2011.

First, using more recent versions of the GEOS-Chem Photochemical Grid model than were available for the 2007 SP, Zhang et al. (2011),[20] Wang et al. (2009),[21] and ICF International[22] (2011) showed that the estimated PRBs were significantly higher than those estimated in the SP. Consequently in the Third Draft of the ISA, EPA used the latest modeling results.

Second, the modeling efforts by Wang et al. and ICF International showed that there were significant contributions to USB from Canadian and Mexican emissions. This is illustrated in Table 1 which shows the results from ICF International. An examination of the spatial distribution of the enhancement reveals the following. For sites in the intermountain West removed from the US-Mexican border, the enhancement is generally small (less than 2 ppb), which is consistent with a relatively minor contribution of Canadian and Mexican sources to the NAB in this region. However, for the Western sites close to the Mexican Border (Joshua Tree, Chiricahua, AZ and Big Bend, TX), the mean enhancement is about 5 ppb. In Big Bend, it increases to 8 ppb. Similarly, for the sites near the Canadian Border (Theodore Roosevelt National Park in North Dakota and Unionville, MI, the mean enhancement is about 5 ppb. The greatest enhancements are seen in Upstate New York (Huntington Forest = 9.0 ppb) and in Maine at Acadia National Park where the mean enhancement is 22.0 ppb. Similar enhancements and geographic patterns were reported by Wang et al.

Site Location	State	USB-PRB	Site Location	State	USB-PRB
Mt. Rainier NP	WA	2.40	Unionville	MI	4.86
Pinnacles NM	CA	1.47	Salamonie Reservoir	IN	3.77
Joshua Tree NM	CA	5.00	Oxford	OH	3.72
Lassen NP	CA	1.13	Perkinstown	WI	3.61
Chiricahua NM	AZ	5.06	Cadiz	KY	2.58
Great Basin NP	NV	2.56	Great Smoky Mtns NP	TN	2.43
Pinedale	WY	1.85	Blackwater NWR	MD	5.57
Rocky Mtn NP	CO	1.91	Acadia NP	ME	21.98
Caddo Valley	AR	2.41	Penn State	PA	5.11
Konza Prairie	KS	2.80	Huntington Wildlife Forest	NY	9.03
Theodore Roosevelt NP	ND	4.64	Sumatra	FL	3.63
Big Bend NP	TX	8.05			

Table 1: GEOs-Chem derived differences between mean daily maximum 8-hour ozone USB and NAB for the 2006, 2007 and 2008 simulations. These values were calculated the values presented in Table 6-3 in ICF Kaiser (2011).

[20] L. Zhang, D. J. Jacob, N.V. Downey, D.A. Wood, D. Blewitt, C.C. Carouge, A. Van donkelaar, D.B.A. Jones, L.T. Murray and Y. Wang, "Improved estimate of the policy-relevant background ozone in the United States using the GEOS-Chem global model with 1/2° × 2/3° horizontal resolution over North America," *Atmos. Environ*. 45:6769-6776 (2011).

[21] H. Wang, H., D.J. Jacob, P. Le Sager, D.G. Streets, R.J.Park, A.B. Giulliland and A. van Donkelaar, (2009), "Surface ozone background in the United States: Canadian and Mexican pollution influences," *Atmos. Environ.* 43:1310-1319.

[22] ICF International (2011), "Modeling for North American Background Concentrations," Contract No. EP-C-09-009, Oct. 28, 2011.

However, just as we have shown that it is inappropriate to use monthly average NAB,[23] it is inappropriate to use mean monthly USB values or mean monthly enhancements. Both ICF and Wang et al. present time series plots for selected sites. In Figure 1, the time series from ICF for Big Bend is reproduced. This clearly shows frequent USB excursions of 50 ppb, a few of 60 ppb and one of 70 ppb. In addition, there are numerous times when the USB is 20 ppb greater than NAB. On three occasions, when the ozone exceeded the 75 ppb NAAQS, the USB ranged from the low 60s to 70 ppb.

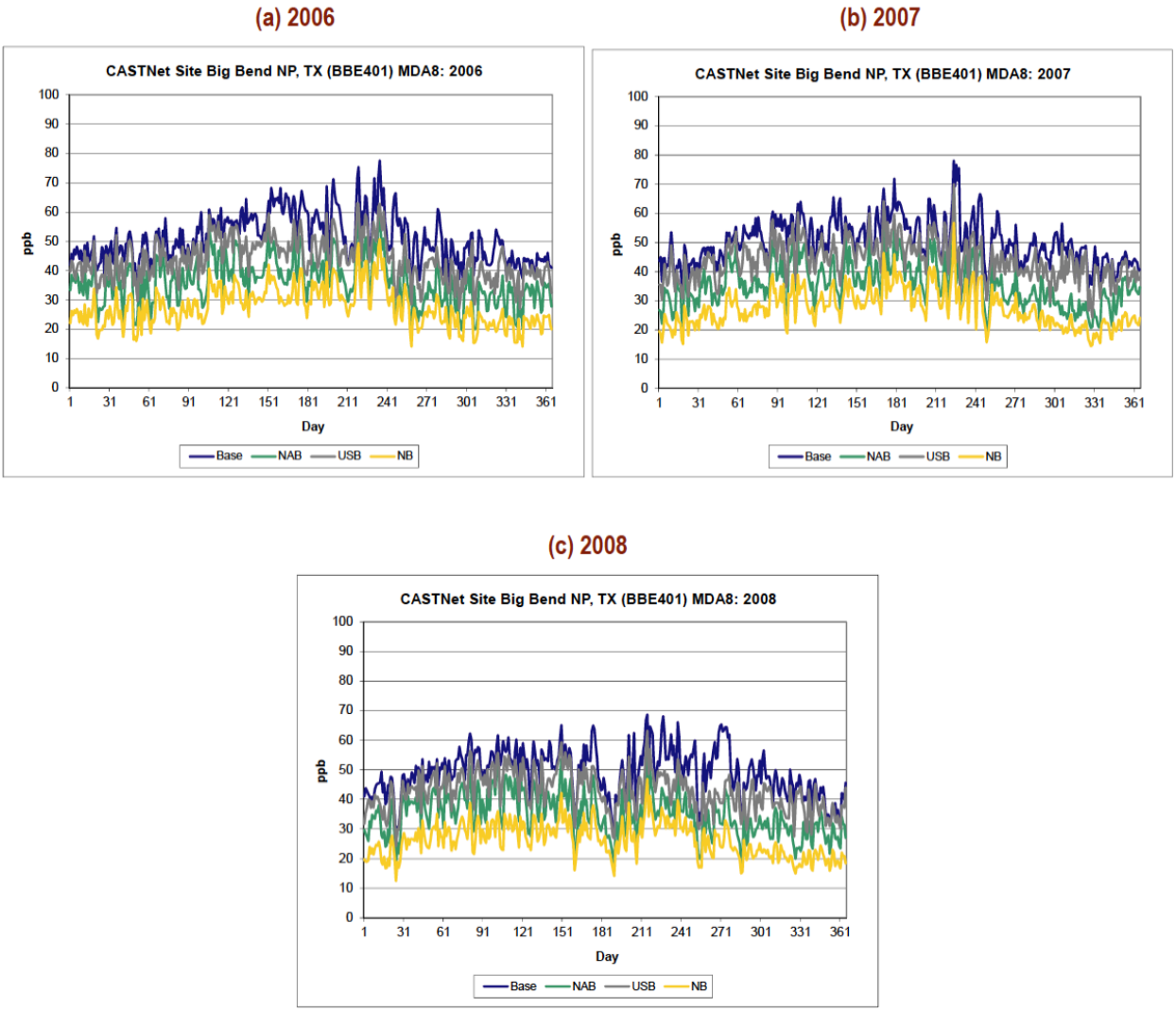


Figure 1: Daily 8-hour maximum ozone concentrations for Big Bend NP, TX. Blue line is the total ozone from the base case, grey is the USB, green is the NAB and the yellow line is the natural background (NB).

[23] Heuss et al., supra note 14, at pp. 7-14.

Figure 2 is the Unionville, MI time series from Wang et al. On several occasions, the USB approached or exceeded 40 ppb and on three occasions, the ozone exceeded the 75 ppb NAAQS while the Canadian contribution exceeded 10 ppb causing exceedances of the 75 ppb NAAQS. Further, there are two occasions when the Canadian enhancement exceeded 30 ppb.

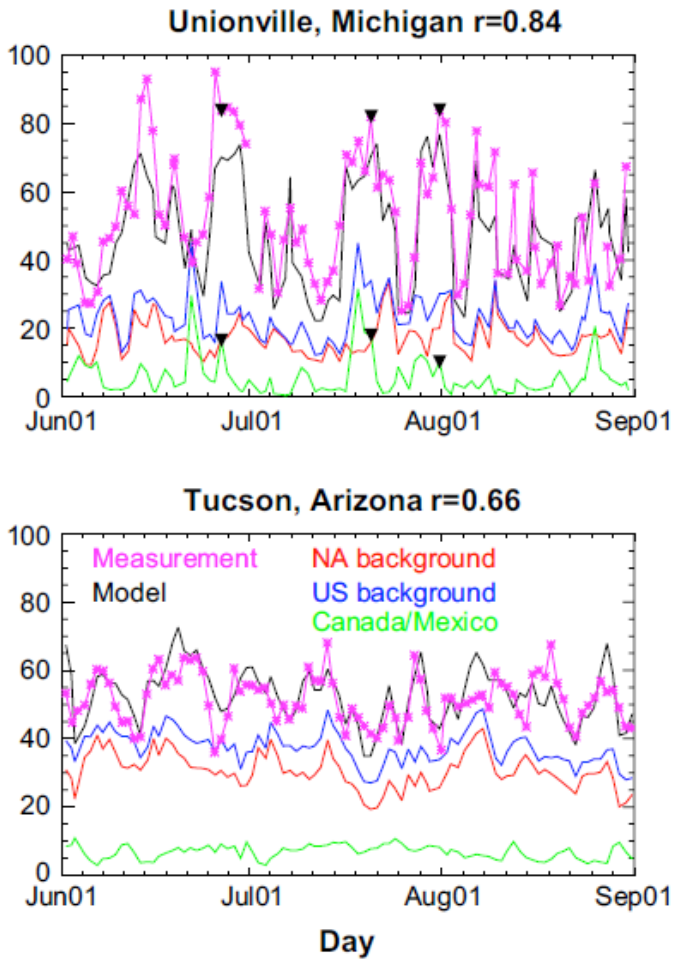


Figure 2: Jun–Aug 2001 time series of daily-8 h-max ozone concentrations at one US site in the Midwest (top) and Southwest (bottom) where Canadian and Mexican influences are particularly strong. Model results (black line) are compared to observations (magenta line with stars). Also shown are the USB (blue line), the NAB (red line), and the Canadian and Mexican pollution enhancement (green line) determined by difference of the USB and the NAB. Black triangles highlight days when observed ozone exceeds 75 ppb and Canadian/Mexican enhancement exceeds 10 ppb. The r is the correlation coefficient between modeled values and observations.

Clearly these modeling exercises provide strong evidence that Canadian and Mexican emissions significantly enhance the USB and this enhancement is occurring on high ozone days and can be a major contributor to exceedances of the 75 ppb NAAQS. As a result, in the Third ISA, EPA decided to also include USB in their analyses.

The third point made from the recent studies is from an application of a higher resolution regional model to estimate NAB. Using the GEOS-Chem model for the boundary conditions, Emery et al. (2012)[24] applied the CAMx model to estimate NAB. They summarize their results:

> In general, CAMx performed better in replicating observations at remote monitoring sites, and performance remained better at higher concentrations. While spring and summer mean PRB predicted by GEOS-Chem ranged 20-45 ppb, CAMx predicted PRB ranged 25-50 ppb and reached well over 60 ppb in the west due to event-oriented phenomena such as stratospheric intrusion and wildfires. CAMx showed a higher correlation between modeled PRB and total observed ozone, which is significant for health risk assessments.

These results indicate that CAMx displayed superior skill over GEOS-Chem and was better able to reproduce the higher concentration days. In addition, it showed that there was a positive correlation between PRB and total ozone which means there is a tendency for higher PRB concentrations when the ozone is higher. Consequently, the importance of including high ozone days in the analyses is underscored. Thus, the daily maximum 8-hour background ozone need to be calculated instead of EPA's practice of using monthly-averaged diurnal profiles of background ozone.

As a consequence of these CAMx results, EPA included CAMx results in their Third Draft ISA and commissioned an internal study[25] to compare the GEOS-Chem and CAMx results. In addition, EPA also began to focus on the high concentration days rather than monthly or seasonal averages.

As indicated from the preceding discussions, EPA is aware of the results of these recent studies and they have begun, in the third draft of the ISA, to include them in their discussions and analyses. However, as we pointed out in our comments on the third draft, much needs to be done before EPA can fully utilize them in their risk and policy assessments. For example, USB estimates have only been made for 2006 using GEOS-Chem. There are no USB estimates available using CAMx and it is not clear that the Agency has the resources or intentions to generate these estimates.

Unfortunately, there appears to be large disconnects between the treatment of background ozone between the third draft of the ISA and the first drafts of the REA and PA. Although the third draft ISA preceded the other two drafts by one to two months, there seems to have been little coordination between the documents' authors on the treatment of background ozone. This is the focus of the next two sections.

---

[24] C. Emery, J. Jung, N. Downey, J. Johnson, M. Jimenez, G. Yarwood and R. Morris, "Regional and global modeling estimates of policy relevant background over the United States." *Atmos. Environ.,* 47:206-217 (2012).

[25] B.H. Henderson, N. Possiel, F. Akhtar and H. Simon, Regional and Seasonal Analysis of North American Background Ozone Estimates from Two Studies . US EPA, Office of Air Quality Planning and Standards memo, August 15, 2012. (available at: http://www.epa.gov/ttn/naaqs/standards/ozone/data/20120814BackgroundOzone.pdf).

3. Background Ozone in the REA

In this section, the discussion will be limited to what background they use. How they use it will be covered in section B.

The REA[26] uses USB estimates from 2006 which they attribute to Zhang et al. (2011).[27] Zhang et al. did calculate USB for 2006, but they do not present the results in the paper. Their only mention of the USB results:

> We find that the US background is on average 1-3 ppbv higher than the North American background, reflecting anthropogenic sources in Canada and Mexico, with little variability except in border regions. Our results for the US background are similar to those reported in the focused GEOS-Chem analysis of H. Wang et al. (2009) and hence we do not discuss them further.

We must therefore assume that Zhang et al. provided EPA with the files that contained the gridded hourly USB estimates. The REA contains the following description of how the Zhang et al. results were used to generate background estimates for 2006 - 2010.

> Background concentrations were estimated from two GEOS-Chem modeling simulations for the model year of 2006: one with zero U.S. anthropogenic emissions (i.e. U.S. background) but with all other anthropogenic and natural emissions globally, and the other with all anthropogenic and biogenic emissions included (i.e. base case) (Zhang et al., 2011). The monitors in each study area were paired with their appropriate GEOS-Chem grid cells, potentially matching multiple monitors to the same cell. The paired hourly U.S. background and base case concentrations were then spatially averaged in the same way as the O_3 monitoring data (as described in 4.3.1.1). Medians by area, month, and hour of the day were calculated from the spatially-averaged U.S. background and base case modeled concentrations, and ratios of the U.S. background to base case concentrations were calculated to provide monthly diurnal profiles of the ratio of U.S. background to total ozone for every month for every area. Next, the U.S. background ratios were multiplied by the respective monitored values in each of the 5 years, 2006-2010, to obtain the U.S. background floor values.[28]

In summary, EPA obtained unpublished, non-peer reviewed USB estimates that are not publically available for 2006 from Zhang et al. and calculated medians by area, month and hour of the day. They then used Zhang's peer-reviewed base case concentrations to estimate monthly diurnal profiles of the ratios which have not been peer-reviewed and are not publically available.

---

[26] U. S. Environmental Protection Agency, supra note 1, at pp. 4-8 - 4-9.
[27] Zhang et al., supra note 20.
[28] U. S. Environmental Protection Agency, supra note 1, at pp. 4-8.

They then used these ratios to produce non-peer reviewed diurnal profiles of the seasonally averaged USB values using 2006 - 2010 ambient measurements for the 12 urban areas selected for their risk assessment. These profiles are shown in Figure 4-3 of the REA which we have reproduced here as Figure 3.

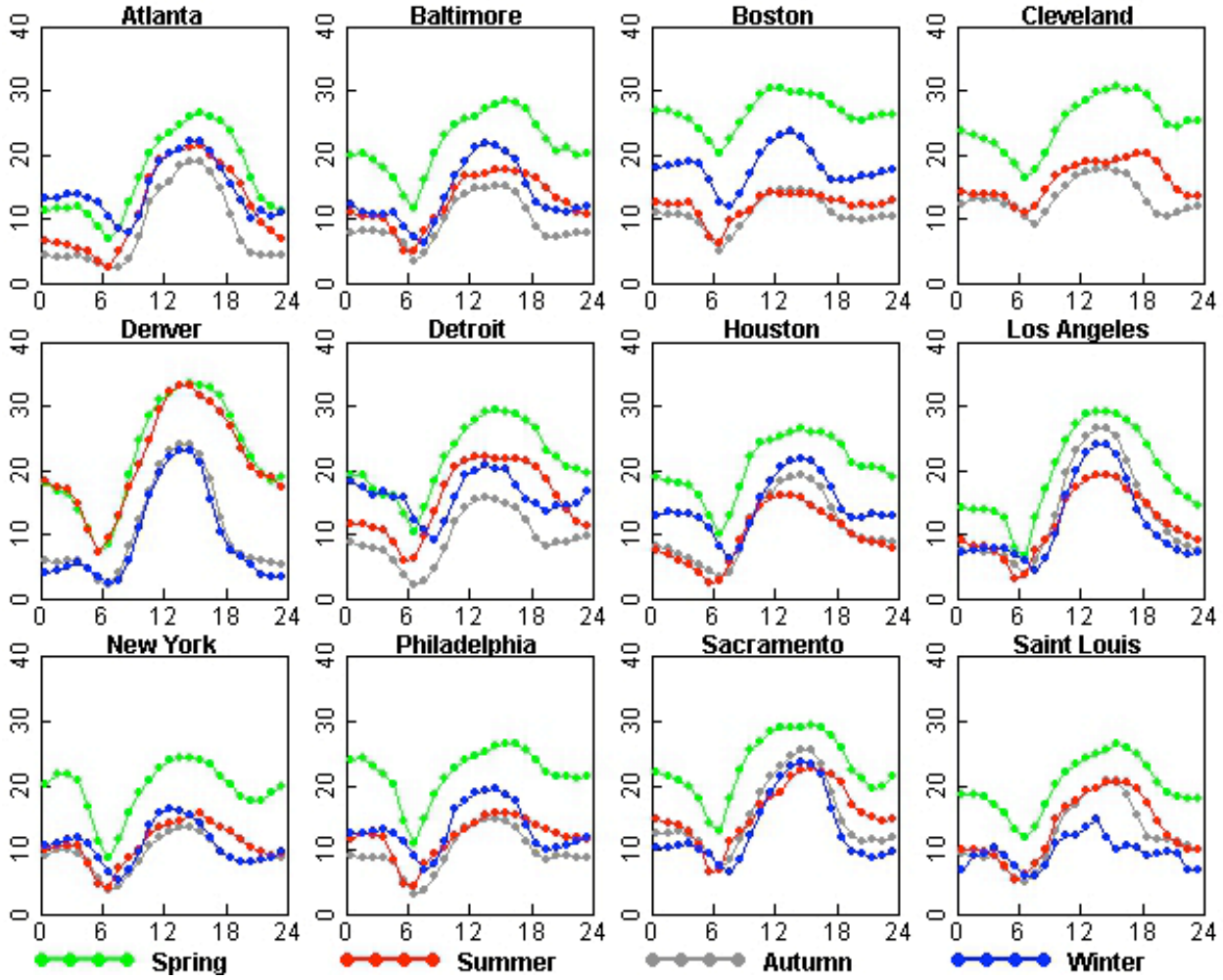
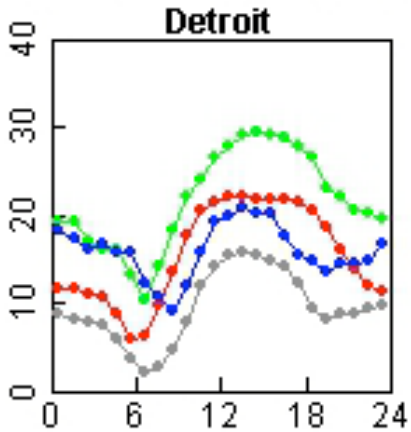


Figure 3: Figure 4-3 in the REA. Diurnal Profiles of Seasonally Averaged U.S Background Floor Values in the Urban Case Study Area**s.** Values shown are 2006-2010 averages, in parts per billion. Seasons were defined as Spring = March – May, Summer = June – August, Autumn = September – November, Winter = December – February. Winter values are missing for Cleveland because no monitoring data were available for that period.

At first glance, the USB values appear to be unrealistically low. To put the USB estimates in perspective, the graph for Detroit was digitized using GraphClick 3.0 from Arizona Software so that the mean MD8H USB values could be computed and compared to the mean MD8H NAB estimates for Detroit that were used in the 2007 SP (see Figure 4). The mean MD8H for USB was 22 ppb in summer and 28 ppb in the spring. This compares to the MD8H NAB in the SP of 22 ppb in the summer and 27 in the spring. Using the latest GEOS-Chem modeling results for

2006 in the third draft of the ISA, EPA estimated the mean maximum daily 8-hour (MD8H) USB for the Detroit area to be 30-35 ppb in the summer and 35-40 ppb in the spring.[29] Consequently, the mean MD8H values for USB presented in the REA and used in EPA's risk assessment are essentially the same as the NAB values used in the 2007 SP for that risk assessment and they are significantly lower than the means presented in the latest ISA. Considering Emory et al.'s conclusions that the superior performance of CAMx compared to GEOS-Chem and the fact that CAMx predicts MD8H values for NAB from CAMx to be about 5 ppb higher than GEOS-Chem's estimates, the range of mean USB MD8Hs in Detroit is more likely 35-40 ppb in the summer and 40-45 ppb in the spring. Thus the mean USB MD8H used in the REA would overestimate the risk reduction that would be realized if the NAAQS were lowered.



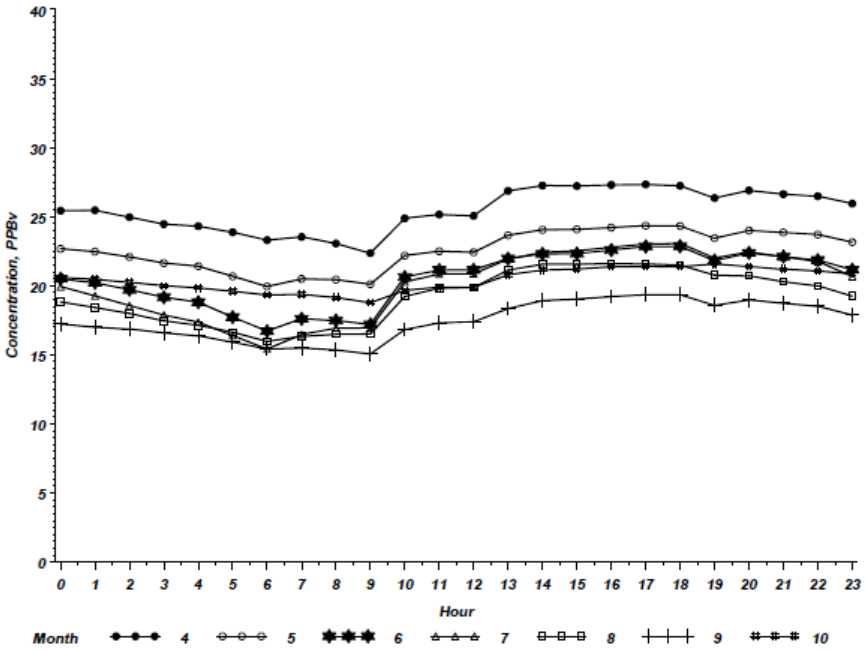


Figure 4: Diurnal ozone profiles for Detroit. Top: USB profile from 2012 REA. Bottom: NAB profile from Appendix 2A in the 2007 SP.

[29] U. S. Environmental Protection Agency, supra note 9, at pp. 3-44.

However, there is another reason why the REA underestimates the impact of USB. The use of seasonal MD8H USB means ignores the positive correlation that was observed by Emory et al. between background and total ozone. As the observed MD8H ozone increases, USB should generally increase. This increase cannot be captured if seasonal means are used to calculate USB. Only daily modeled estimates of MD8H USB using CAMx would reproduce this effect.

In summary, the non-peer reviewed estimates of USB that are contained in the REA and thus used in calculating the risk estimates are gross underestimates. EPA underestimates the mean USB by 15-20 ppb in the spring and summer and the underestimates could even be greater on days with the higher USB contributions. Consequently, EPA risk reduction estimates will be inflated.

4. Background Ozone in the PA

The discussion of background ozone in the PA[30] is quite limited and is based mainly on NAB discussions presented in the ISA and in Henderson et al. (2012).[31] Only one paragraph discusses USB:

> GEOS-Chem results for seasonal mean MDA8 USB and NAB concentrations suggest that USB concentrations are on average 1-3 ppb higher than NAB background, reflecting the influence of anthropogenic sources in Canada and Mexico. Very little variability was found in these concentrations, except in areas in the U.S. that bordered Canada and Mexico, where international transport from these two countries plays a greater role in contributing to O_3 background concentrations (US EPA, 2012a, Figure 3-9). These results were similar to those reported by Wang et al. (2009).[32]

It is of interest to note that the PA refers to the USB values in Figure 3-9 in the ISA, which is the same Figure we cite in footnote 29 as evidence that the REA USB estimates are significantly underestimated.

However, the background ozone summary section (1.3.4.3 on page 1-22) tends to downplay the importance of USB or NAB:

> While numerous large urban areas in the U.S. experience high ambient O_3 concentrations during the warm season, recent modeling efforts indicate that anthropogenic emission sources are the dominant contributor to these ambient concentrations (US EPA, 2012a, section 3.4.3 and Henderson et al., 2012). In the Southeast, Northeast, and North Central, background concentrations were lower in the summer (than in the spring) when measured O_3 concentrations are usually the highest and the 4th highest

[30] U. S. Environmental Protection Agency, supra note 2, at pp. 1-16 - 1-22.
[31] Henderson et al., supra note 25.
[32] U. S. Environmental Protection Agency, supra note 2, at pp. 1-16 - 1-17.

> MDA8 values usually occur. In addition, the GEOS-Chem and CAMx model 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 and, therefore, that anthropogenic sources within the U.S. are largely responsible for 4th highest 8-hour daily maximum O_3 concentrations. In areas where background O_3 is highest, such as the western U.S. and at higher elevation sites, the sources contributing to high background concentrations have been identified as wildfires, stratospheric intrusions, and intercontinental transport (US EPA, 2012a, section 3.4.3). As noted above, EPA has policies that allow exclusion of air quality monitoring data affected by these types of events.[33]

The first two sentences suggest a minor role for USB or NAB especially during the summer. However, the data presented in Table 3-1 of the ISA (Table 2 in this report) disputes that claim. Clearly during the summer the NAB is the *largest* source of ozone comprising on average more than 50% of the of the observed ozone.

Table 3-1 Comparison of seasonal mean MDA8 ozone concentrations simulated by the GEOS-Chem and CAMx base case models for 2006, with measurements at CASTNET sites.

Region	CASTNET		GEOS-Chem		CAMx	
	Spring	Summer	Spring	Summer	Spring	Summer
California (5)[a]	58 ± 12[b]	69 ± 14	52 ± 11; 0.52[c] 38 ± 7[d]	66 ± 18; 0.22 37 ± 9	50 ± 10; 0.50 39 ± 6	66 ± 13; 0.30 42 ± 6
West (14)	54 ± 9	55 ± 11	53 ± 7; 0.30 42 ± 6	55 ± 11; 0.12 40 ± 9	49 ± 8; 0.39 40 ± 7	57 ± 10; 0.33 41 ± 8
North Central (6)	47 ± 10	50 ± 12	47 ± 8; 0.52 33 ± 6	51 ± 14; 0.44 27 ± 7	45 ± 11; 0.63 30 ± 6	54 ± 13; 0.48 31 ± 5
Northeast (5)	48 ± 10	45 ± 14	45 ± 7; 0.44 33 ± 7	45 ± 13; 0.47 24 ± 7	46 ± 11; 0.53 30 ± 5	53 ± 14; 0.54 27 ± 6
Southeast (9)	52 ± 11	52 ± 16	51 ± 7; 0.42 32 ± 7	54 ± 9; 0.21 29 ± 10	54 ± 9; 0.56 33 ± 6	61 ± 12; 0.45 30 ± 6

[a]Values in parentheses after each region name refer to the number of sites.

[b]Shown are seasonal (spring, summer) mean MDA8 O_3 concentrations (ppb ± standard deviation);

[c]Shown are mean R^2 of all model-measurement pairs at individual CASTNET sites.

[d]North American (NA) background seasonal mean MDA8 O_3 concentrations (ppb ± standard deviation) are shown beneath the base case seasonal means.

Source: Data from Zhang et al. (2011) for GEOS-Chem and Emery et al. (2012) for CAMx.

Table 2: Comparison of seasonal mean MD8H measurements, model estimates and model estimates of NAB.

The third sentence states that background ozone is "not dramatically higher than typical seasonal average background concentrations" on high ozone days. As shown in Figures 1 and 2 by Emory

[33] U. S. Environmental Protection Agency, supra note 2, at p. 1-22.

et al. and Wang et al., this is not always the case and, in general, NAB tends to go up with total ozone. Similar results to Emory et al.'s have been reported by Downey et al. (2011)[34] using GEOS-Chem as well.

The last two sentences imply that areas that experience exceedances that can be blamed on high background due to wildfires, stratospheric intrusions or intercontinental transport need not be concerned because EPA has policies in place to exclude such occurrences. This ignores the history of such occurrences where only the most blatant incidences have been excluded and the high bar that EPA requires the states to hurdle.

Based on the tone and content of this summary paragraph, it appears that EPA is trying to minimize the importance of USB to the policy makers. This is unfortunate because it downplays the difficulties the states will face without properly appreciating the significant role of USB in the development of control strategies. In addition, it masks the significant contribution that USB makes to the risk estimates made by EPA.

B. EPA's Use of Background Ozone in the Risk and Policy Assessments

In this section, we will focus on how EPA uses background ozone to estimate risk. The approach used in the current REA differs significantly from the way EPA calculated risks in their previous risk assessments.

1. The 2007 SP

To calculate the risk reduction that would occur if emissions were reduced so that the NAAQS was just being achieved, EPA employs the "quadratic rollback" method which they describe:

> The "quadratic rollback" method was used in the previous O_3 NAAQS review to adjust ambient O_3 concentrations to simulate minimally meeting current and alternative standards (U.S. EPA, 2007)[35]. As the name implies, quadratic rollback uses a quadratic equation to reduce high concentrations at a greater rate than low concentrations. The intent is to simulate reductions in O_3 resulting from unspecified reductions in precursor emissions, without greatly affecting concentrations near ambient background levels.[36]

The quadratic rollback was used to simulate reductions in areas that did not meet the NAAQS. Hourly concentrations were reduced so that the area's design value was exactly equal to the NAAQS. Concentrations at other monitors were similarly reduced using the quadratic rollback coefficients calculated at the highest monitor. They then would apply the coefficients to the "as is" measured ozone concentration distribution to compute a new distribution that corresponded to meeting the NAAQS.

---

[34] Downey et al., supra note 19.
[35] U. S. Environmental Protection Agency, supra note 6.
[36] U. S. Environmental Protection Agency, supra note 2, at 4-7.

They then applied what they considered to be the appropriate no-threshold dose-response functions to compute the relative reduction in risk that would be achieved with the new distribution meeting the NAAQS. However, **they only considered the risk above background concentrations,** which, for the 2007 review, was defined as PRB (the bottom graph in Figure 4 are the PRB values EPA used for Detroit). Any concentration that was at or below the PRB was excluded from the risk calculation because EPA considered this to be residual risk that could not be reduced by additional regulations or international agreements.

AIR, Inc. agreed that the risk during periods with concentrations at or below background should be excluded because of the reasons EPA gave - it was beyond our control to reduce. However, as articulated in the previous sections, AIR objected to using PRB instead of USB and AIR demonstrated that the PRB values that EPA used were unrealistically low.

2. The Present Review

In this review EPA again uses the quadratic rollback:[37]

> In this review, quadratic rollback was used to simulate reductions in O_3 concentrations in areas which failed to meet EPA's current O_3 NAAQS of 0.075 ppm (75 ppb). Hourly O_3 concentrations were reduced so that the highest design value in each area was exactly 75 ppb, the highest value meeting the NAAQS. Concentrations at the remaining monitors in each area were similarly reduced using the quadratic rollback coefficients calculated at the highest monitor. Quadratic rollback was performed independently within each area for two design value periods, 2006-2008 and 2008-2010. In some of the 12 urban areas, the monitor with the highest design value was not within the area boundaries chosen to match the study areas in Zanobetti & Schwartz (2008).[38] In these cases, the high monitor was included in the quadratic rollback, and the ozone concentrations at the monitors within the Zanobetti & Schwartz (2008) study area were similarly reduced. In this way, while the high monitor outside of the study area would have been simulated to have a design value of 75 ppb to just meet the standard, the design value at the monitors within the study area would have been simulated to have design values below 75 ppb.
>
> To avoid reducing O_3 concentrations below background levels, background "floor" values were set defining minimum values beyond which quadratic rollback would not be applied.

Up to this point, the methodology is similar to that employed in the 2007 SP. However, as EPA explains below, a different approach was used from here on:[39]

---

[37] Ibid.

[38] A. Zanobetti and J. Schwartz, "Mortality displacement in the association of ozone with mortality: An analysis of 48 cities in the United States." *Am. J. Resp. Crit. Care Med.*, 177:184-189 (2008).

[39] U. S. Environmental Protection Agency, supra note 2, at 4-34 - 4-35.

For example, in contrast to the approach used in the last review, the first draft REA has estimated total risks attributable to O_3 exposure, *not risks in excess of background concentrations*. In taking this approach, the REA noted the advice of CASAC members, who recommended in the last review that EPA move away from using background in calculating risks (Henderson, 2007).[40] This approach recognizes that health risks result from O_3 exposures, regardless of the source of the O_3.

In estimating total O_3-related health risks, the REA concluded that the approach most consistent with the statistical models reported in the epidemiological studies is to apply the concentration-response functions to all ozone concentrations *down to zero*. However, consistent with the conclusions of the ISA that the available evidence indicates less certainty in the shape of the concentration-response curve at the lower end of the distribution of ambient O_3 concentrations, the REA also recognized that confidence in the nature of the concentration response function and the magnitude of the risks associated with very low concentrations of ozone is reduced because there are few ozone measurements at the lowest levels in many of the urban areas included in the studies. Specifically, the REA noted that estimates of risk associated with O_3 concentrations below the lowest measured level (LML) for the underlying epidemiological study would be associated with reduced confidence since these estimates involve applying the concentration-response function outside of the range of data used in its derivation. In light of this, the REA has characterized mortality risks in excess of lowest measured O_3 concentrations as well as total risks associated with O_3 concentrations down to zero (US EPA, 2012b, sections 7.3.3 and 8.1.1.4).[41] In considering these different approaches, the REA concluded that the two sets of estimates provide a reasonable bound on estimated total risks, reflecting uncertainties about the concentration-response functions below the lowest ozone concentrations evaluated in the studies.

In the remainder of this section, we consider the first draft REA estimates of O_3-related health risks within the context of considering the adequacy of the current standard. Specifically, we consider risk estimates for all-cause mortality and respiratory morbidity, which includes respiratory-related hospital admissions, emergency department visits, and symptoms (Emphasis added).

---

[40] R. Henderson, Letter from CASAC Chairman Rogene Henderson to EPA Administrator Stephen Johnson. March 26, 2007, EPA-CASAC-07-002 (2008). Available at: http://yosemite.epa.gov/sab/sabproduct.nsf/FE915E916333D776852572AC007397B5/$File/casac-07-002.pdf.

[41] U. S. Environmental Protection Agency, supra note 9.

Consequently, EPA has decided to ignore the fact that risk due to background ozone cannot be reduced through control strategies and instead includes that risk in their risk estimates. This of course has the effect of further inflating their risk estimates. The risk estimates down to zero further maximizes the estimates. The reason they use to justify ignoring background is that it was recommended by CASAC in the 2007 memo from Henderson to Administrator Johnson. One CASAC Panel member, Dr. Fred Miller, who serves on both the present review and the 2007 review, disagrees as he says EPA Staff is misinterpreting what Henderson wrote. In his comments on the First Draft of the PA, Dr. Miller wrote:[42]

> This reviewer does not agree with the interpretation that the staff have taken relative to the suggestion by CASAC during the previous NAAQS review cycle where they state starting on page 4-34 "In taking this approach, the REA noted CASAC members, who recommended in the last review that EPA move away from using background in calculating risks (Henderson, 2007)". This reviewer believes staff misinterpreted CASAC's advice. The full paragraph from the Henderson (2007) memo is stated below.
>
>> Finally, with respect to policy-relevant background (PRB), the Ozone Panel wishes to point out that the Final Ozone Staff Paper does not provide a sufficient base of evidence from the peer-reviewed literature to suggest that the current approach to determining a PRB is the best method to make this estimation. One reason is that part of the PRB is not controllable by EPA. It would require international cooperation beyond the bounds of North America. A better scientific understanding of the PRB and its relationship to intercontinental transport of air pollutants could serve as the basis for a more concerted effort to control its growth and preserve the gains in air quality achieved by control efforts within the U.S. In any case, there is no apparent need to define PRP in the context of establishing a health-based (primary) ozone NAAQS. The effects of inhaled ozone on decreases in respiratory function have been seen in healthy children exposed to ozone within ambient air mixtures in summer camps (1–6). Furthermore, the concentration- response functions above 40 ppb are either linear, or indistinguishable from linear. Thus, PRB is irrelevant to the discussion of where along the concentration-response function a NAAQS with an 8-hour

[42] F. Miller, Comments from Members of the CASAC Ozone Review Panel for Discussion at the September 11-13, 2012 Meeting, pp. 32-33. Available at: http://yosemite.epa.gov/sab/sabproduct.nsf/BAF7E636BBCA39E785257A7600464707/$File/Preliminary+Individual+Comments+on+PA+9-11-12.pdf.

> averaging time that provides enhanced public health protection should be.[43]
>
> What CASAC was conveying was concerns about the state of knowledge of PRB levels at that time and that selecting the range to consider for setting the O_3 NAAQS based on the scientific evidence for health effects did not need to have the PRB level enter into the process. However, from a science policy and risk management judgment perspective, the Administrator must be made aware of the portion of the total risk for a given health endpoint that exists over which EPA regulatory action would not have any control – namely the portion of total risk from zero to the policy background level in order that she/he can execute their responsibilities under the Clean Air Act to set primary NAAQS that are neither more nor less stringent than necessary to protect public health with an adequate margin of safety.
>
> Since the last review cycle, a great deal more of information about background levels of O_3 excluding anthropogenic sources has become available from a combination of measurements and atmospheric models. Moreover, there is a better understanding of how these levels vary during seasons of the year in different geographical regions. This translates into increased confidence in the Agency being able to incorporate adequately region specific background levels into their risk assessments and account for different cities in epidemiology studies having different background O_3 levels and, therefore, different reductions in risk values when considering alternative standard levels. *Failure to do such invites litigation against the Administrator and the Agency* (Emphasis added).

3. Discussion

When a no-threshold dose-response risk function is used, a good proportion of the calculated risk occurs during hours when the ozone concentrations are at or below USB concentrations. This proportion will become larger as the quadratic rollback is applied to even lower standards.

EPA attempts to justify their approach by saying it is what CASAC wanted in 2007. However, one CASAC member has conveyed the correct intent of CASAC's 2007 message: "t(T)his translates into increased confidence in the Agency being able to incorporate adequately region specific background levels into their risk assessments and account for different cities in epidemiology studies having different background O_3 levels and, therefore, different reductions in risk values when considering alternative standard levels." AIR also agrees with the CASAC member's assessment in his final sentence: "f(F)ailure to do such invites litigation against the Administrator and the Agency."

C. Summary

[43] Henderson, supra note 40, at 2-3.

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 assessment of risk which is then used to inform policy decisions that need to be made concerning the form, averaging time and level of the NAAQS.

In the current review, EPA uses unrealistically low values for the USB. The mean diurnal estimates of USB that EPA presents in the REA underestimate the USB MD8H values by 15-20 ppb in the spring and summer and could even be greater on days with the higher USB contributions. Consequently, the risk reductions calculated from these USBs are inflated.

In addition, they use a methodology in their assessment of risk that obscures the important contribution that background ozone makes to the risk estimates calculated by EPA. The contribution of risks occurring at concentrations of ozone at or below background becomes more important as lower NAAQS are considered. The methodology EPA uses to calculate risk, assumes no threshold concentration for health effects and assumes that exposure to concentrations of ozone at or below background levels pose a real threat to human health. These assumptions inflate the estimated health risks and the estimated health risk reductions when more stringent NAAQS are considered.

II. Comments on the Draft REA

The REA provides preliminary quantitative estimates of the health risk from ozone. It builds upon and uses much the same methodology as that used in the last review completed in 2008. Although some updates to the methodology are indicated in the REA, such as additions to the Consolidated Human Activity Database (CHAD), all changes from the last review included in the second draft REA should be fully documented. In addition, in the process of evaluating the Air Pollution Exposure Model (APEX), AIR identified a few issues with the FORTRAN coding that may cause problems when the model is applied by either EPA or outside parties. These issues are discussed in Appendix 2.

The REA list a number of goals and the document is organized around the efforts to meet these goals. AIR has comments on each of the 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 not directly a measure of risk of adverse effects or risk to public health. Although the benchmarks chosen -- 8-hour exposures of >0.060 ppm, >0.070 ppm and >0.080 ppm -- 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 at any of these 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 the risk,

all these factors need to be taken into account and this is what APEX attempts to do. Because of the importance of exercise, the portions of the model that simulate activity and ventilation rate need special scrutiny.

AIR has identified three ways in which the estimates of benchmark exposures in Chapter 5 the draft REA are biased high. 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] 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.[45] This was an important oversight because the upper percentiles of ventilation rate are responsible for the exposures that cause the perceived risk. In the comparison of the APEX modeled values with the measured ventilation rates from Brochu et al. (2006),[46] the model over-predicted mean daily ventilation rates for persons below age 11 and over age 40. More importantly, the model had a much higher standard deviation at all ages.

This suggests that the upper percentiles of ventilation rates in the model are substantially above those measured 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. Figure 5 shows that the APEX model EPA used in the prior risk assessment significantly overestimates the breathing rates of male children, particularly for the upper tails of the distribution that are responsible for the exposures of concern evaluated by the Agency. The data underlying these distributions (means and standard deviations) come from Table 25 in the 2007 Langstaff Memorandum on uncertainty in the exposure model. In fact, of the 16 comparisons in Table 25, for eight age groupings each of males and females, 15 had substantially higher modeled ventilation rates compared to the data reported by Brochu et al., 2006 at the upper end of the distribution.

---

[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] Ibid., at pp. 52.
[46] 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).

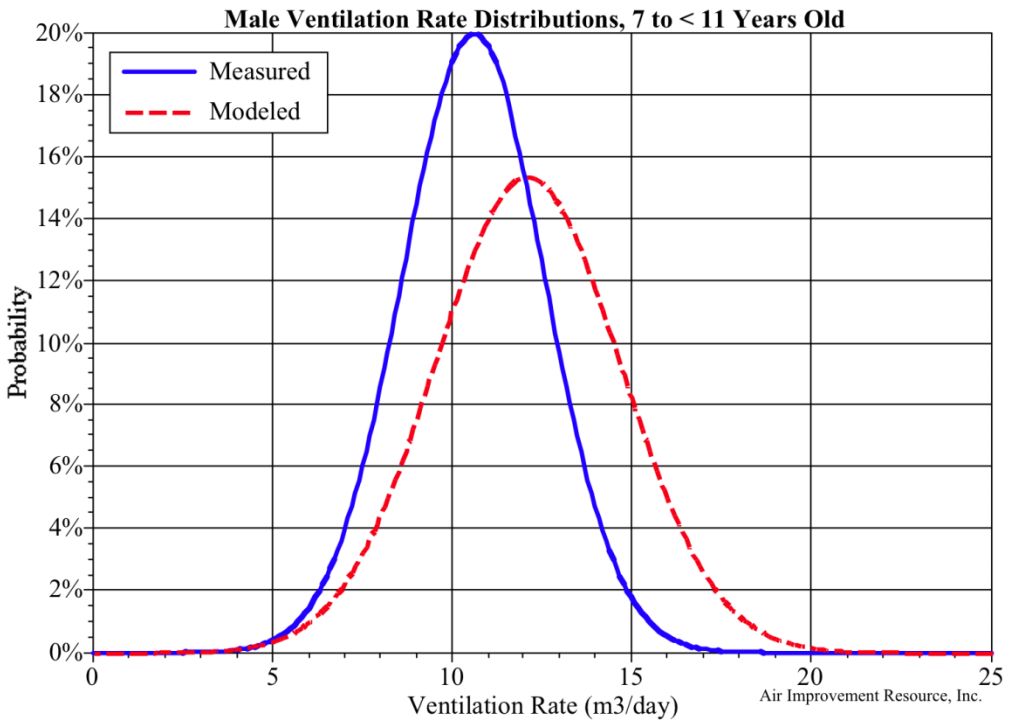


Figure 5: Comparison of measured vs. modeled daily ventilation rates for 7- to 10-year old boys.

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.[47]

The limited discussion of the methodology in the current draft refers to METs (metabolic equivalents of work) values derived from Ainsworth, et al., 1993 using a formulation described in McCurdy, 2000. There have now been two updates to the Ainsworth et al. Compendium of Physical Activities. The latest was published in 2011. The REA indicates that the second draft will include the use of updated METs data. Because of the acknowledged upward bias in the ventilation rate data used in APEX and because of the importance of the methodology for assigning ventilation rates to the estimated risk, the second draft REA should include a detailed discussion of the current methodology and data involved including comparisons to the extremes of real-world data.

A second way the counts of benchmark exposures are biased high relates to how EPA defines moderate or greater exercise over 8 hours. The REA follows the approach begun in 1996 of defining Equivalent Ventilation Rates (EVRs) between 13 and 27 as moderate.[48] The counts in

[47] 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.
[48] REA, supra note 1, at pp. 5-16.

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. Ted Johnson 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.[49] Johnson generated EVR sequences for 4678 subjects in the 18-35 age range using algorithms similar to those in the APEX-ozone model available in 2007. The results are shown in the following Figures 6 and 7 for both asthmatics and non-asthmatics. The distributions of mean EVR, maximum 2-hour EVR and maximum 8-hour EVR (EVR480) are shown. Note that APEX accumulates headcounts 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 headcounts overestimate the number of subjects at potential risk and the resulting risks calculated in Chapter 6 are unreasonably high.

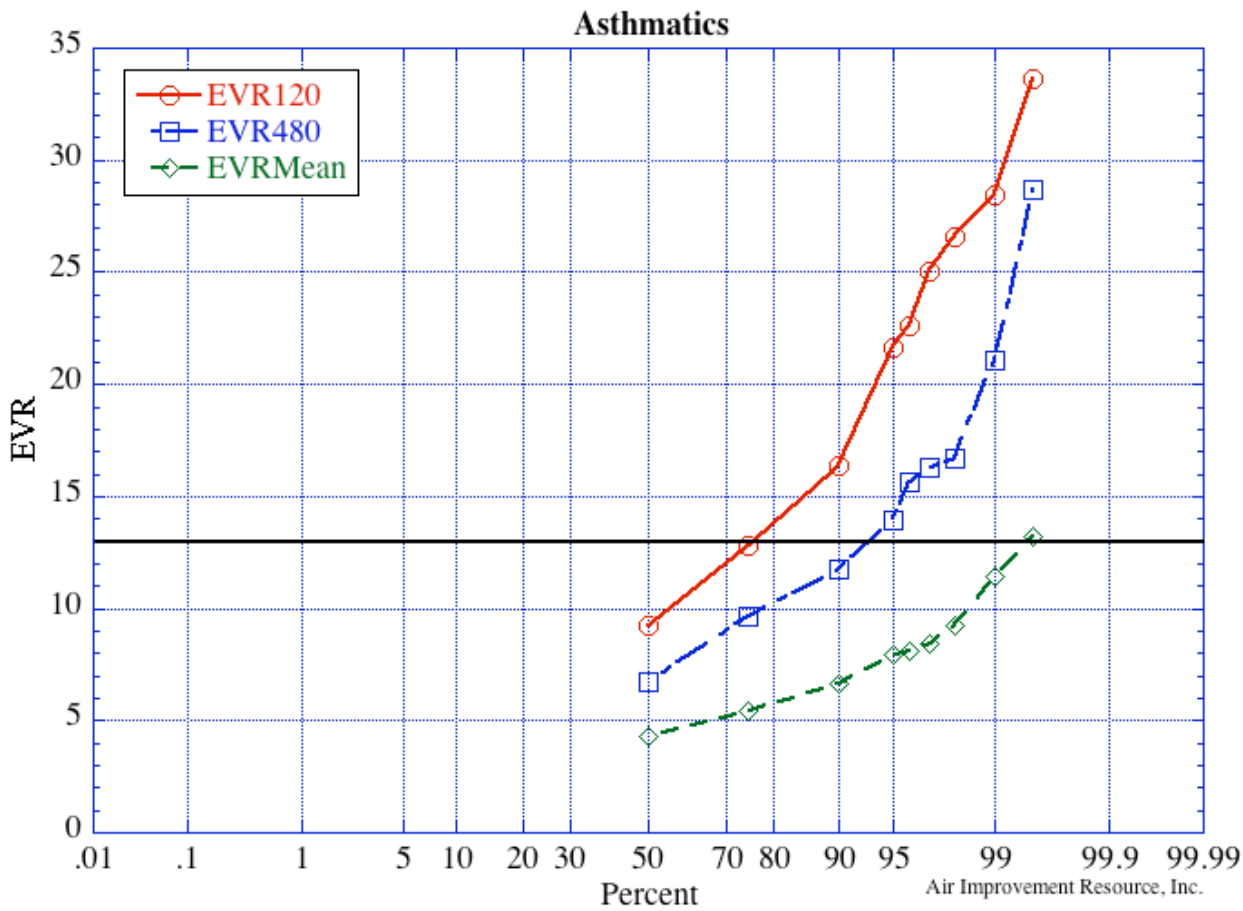


Figure 6 Distribution of EVRs calculated by the APEX algorithm for asthmatics

[49] 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.

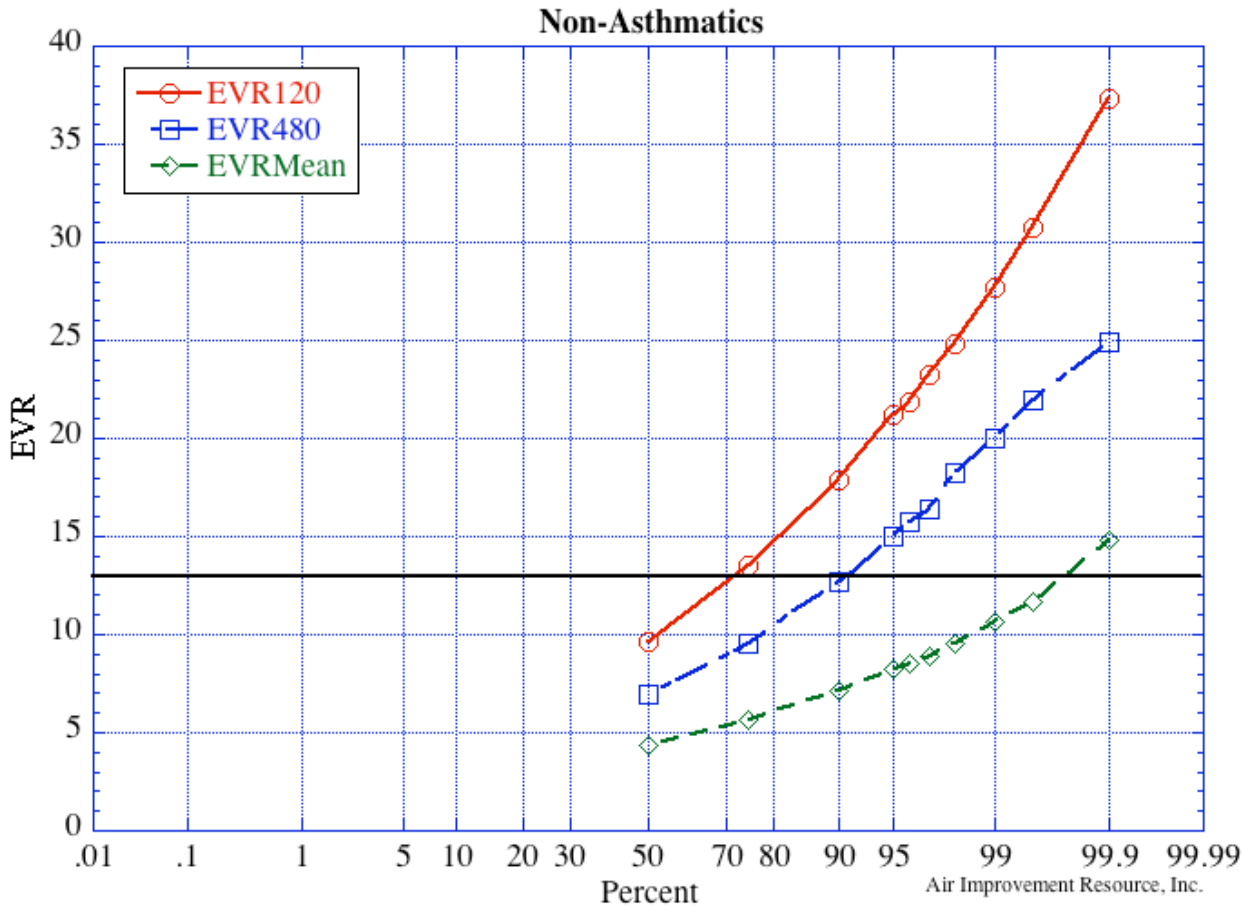


Figure 7 Distribution of EVRs calculated by the APEX algorithm for non-asthmatics

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[50] 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

[50] W. Linn, C. Spicer, and J. Hackney, "Activity patterns in ozone-exposed construction workers," *J. Occup. Med. Toxicol.*, 2, 1-14 (1993).

6.6 hour exposure protocols.[51] 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.

Third, the REA should acknowledge 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)[52] measured the increment between ozone at 2 and 10 meters and reported an average 13 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 0.06 ppm and higher was halved with a 10 percent vertical correction.[53] 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 REA include a discussion of the difference between ozone at person height and at measurement height and the sensitivity to this bias should be evaluated in the next draft.

The presentation of the output of the headcount analysis in the REA is misleading and not directly relevant to public health. The REA 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 as the second metric, counts of person –occurrences over the entire group and ozone season. Single occurrence 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 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 draft REA focuses only on the

---

[51] 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.

[52] 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.

[53] 2007 SP, supra note x, at pp. 7-46 and 7-47.

first metric. There are 19 figures and 4 large tables in Chapter 5 giving the results for the first metric and no presentation at all for the second metric. This must be remedied in the second draft.

Although the full documentation of the APEX runs in the draft REA is not available, there is information in the sample applications for APEX 4.5 on the EPA website that can be used to show the relevance of the second metric. The files contain output of APEX simulations for 5-18 year old children for 2006 ozone levels in Denver and for 2008 ozone levels in Los Angeles. Although only 60 activity profiles were simulated, the results for metric one correspond closely to the Denver results for children in Figures 5-1 and 5-2 in the REA. The results for metric two are presented in Tables denoted PERSONDAYS, DM8H, ALL, ALL and PERSONDAYS, DM8H, MOD, ALL. The Denver simulation estimates 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 predicts that only 0.004 or 0.4 percent of the children's 8-hour exposures are 0.06 ppm or greater. In 2006, the ozone design value for Denver was 0.090 ppm, which is greater than the current ozone standard of 0.075 ppm. For all exposures at 13 EVR or greater, APEX predicts that only 0.0027 or 0.27 % occur at 8-hour exposures of 0.06 ppm or greater. For the cutpoint of 0.070 ppm, the portion of maximum 8-hour exposures with EVR of 13 or greater is 0.00057 or 0.057 %. Thus, in the 2006 base case, the vast majority of children's exposures are below the level of any concern. This is shown in Table 3.

Table 3 - Denver 2006 Base Case

8-h Ozone w/Exercise	Percent Persons	Percent Person-Days
>0.06	30	0.27
>0.07	10	0.057

The Los Angeles sample application provides a similar comparison. The % persons data in Table 4 comes from Figures 5-7 and 5-8 of the REA and the % person-days comes from the sample APEX application.

Table 4 - Los Angeles 2008 Base Case

8-h Ozone w/Exercise	Percent Persons	Percent Person-Days
>0.06	35	0.4
>0.07	18	0.1

In both these sample applications, the vast majority of children's exposures are below those of concern based on the clinical studies. 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 second draft should include presentation of both persons and person-days exposure results.

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

decrements

The second listed goal in the REA is to provide estimates of the number of people with various FEV1 decrements. 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 in lung function at moderate to severe levels of exertion.

However, the calculations presented in Chapter 6 are necessary but not sufficient to estimate the risk of adverse effects since they do not include estimates of lung function decrements accompanied by respiratory symptoms, as the American Thoracic Society Guidelines recommend.[54] As shown in Table 6-7, the risks of FEV1 decrements in the draft REA are similar to those estimated in the last review.

Chapter 6 describes two approaches to estimating the risk. In the first, probabilistic exposure-response functions are applied to the APEX-estimated population distribution of 8-hour maximum exposures at or above moderate exertion to estimate the number of persons expected to experience lung function decrements. This is the approach used in the 2007 review. Although the REA indicates that 8-hour exposures in the EVR range of 13-27 was selected to match the EVR for the group of subjects in the controlled human exposure studies that were the basis for the exposure-response functions used in this portion of the risk assessment, the distribution plots in Figures 6 and 7 indicate that the choice of 13 for the lower limit results in a dramatic overstatement of the number of exposures with EVRs comparable to those used in the clinical studies. The original choice of the range of 13 to 27 to bracket the clinical studies carried out at 20 may appear reasonable until one considers the strenuous nature of the exercise at 20 and the range of actual 8-hour VRs and EVRs in the populations of interest. AIR urges EPA to evaluate the sensitivity of the exposure and risk calculations to different binning assumptions.

The second approach uses a McDonnell-Stewart-Smith FEV_1 model based on 2007 and 2010 publications. That model uses the time-series of O_3 exposure and corresponding ventilation rates for each APEX simulated individual to estimate their personal time-series of FEV_1 reductions, selecting the daily maximum reduction for each person. The REA indicates that EPA will be updating the C-R functions using both approaches for the second draft. The 2007 and 2010 analyses have been superseded by a new McDonnell et al., 2012[55] concentration-response function model. AIR urges EPA to consider the recent McDonnell et al., 2012 and Schelegle et al., 2012[56] analyses as it evaluates alternative C-R functions for use in the risk assessment. Both studies demonstrate that the first physiological effects of ozone have a threshold.

It is also important to discuss how the FEV1 decrement results relate to adversity. Single small

---

[54] "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).

[55] 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).

[56] E. S. Schelegle, W. C. Adams, W. F. Walby, and M. S. Marion, "Modelling of individual subject ozone exposure response kinetics," *Inhal. Toxicol*., 24, 401–415 (2012).

transient FEV1 decrements without symptoms have not been considered adverse in prior reviews for either asthmatics or normal subjects.[57] All the data presented in Chapter 6 relates to the percent of subjects experiencing one or more FEV1 decrements an ozone season. This statistic is not particularly informative. 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.

For example, the comparisons in Tables 3 and 4 above for ozone exposures can be extended to include FEV1 decrements. Tables 5 and 6 below provides a comparison for the 2006 Denver and 2008 Los Angeles base cases of ozone exposures and FEV1 decrements in terms of metric one (% of persons experiencing the given factor) and metric two (% of total person-days experiencing the given factor). The % persons data in Table 5 for Denver come from Figures 5-1 and 6-9 of the REA. The % persons data in Table 6 for Los Angeles come from Figures 5-7 and 6-10 of the REA. The % person-days >0.06 come from the APEX sample calculation discussed above, and the % person-days for various FEV1 increments were estimated assuming the same portion as for percent days.

Table 5 – Denver 2006 Base Case

Percent Persons				**Percent Person-Days**			
8-h>06	FEV>10	FEV>15	FEV>20	8-h>06	FEV>10	FEV>15	FEV>20
30	7.5	2.5	0.6	0.27	0.068	0.023	0.005

Table 6 – Los Angeles 2008 Base Case

Percent Persons				**Percent Person-Days**			
8-h>06	FEV>10	FEV>15	FEV>20	8-h>06	FEV>10	FEV>15	FEV>20
35	9	3.5	1	0.4	0.10	0.04	0.01

Based on this comparison, the percent of person-days with children experiencing FEV1 decrements in Denver and Los Angeles under current air quality is extremely small. Thus using the REA methodology, current air quality is very protective of public health. Attainment of the current standard would reduce these already extremely small risks substantially. To provide a more complete perspective on the public health impact of the current and alternative standards, the second draft should include presentation of both persons and person-days results.

C. Estimates of the potential magnitude of mortality and morbidity risks based on epidemiological studies

The third goal is to provide estimates of the potential magnitude of premature mortality and/or selected morbidity health effects.

The REA points out:

[57] 1996 SP, supra note x, at pp. 62-72.

> EPA also acknowledged that at the time of the previous review there were considerable uncertainties surrounding estimates of O3 C-R coefficients and the shape for concentration-response relationships and whether or not a population threshold or non-linear relationship exists within the range of concentrations examined in the epidemiological studies.

The REA further notes:

> We have identified multiple options for specifying the concentration-response functions for particular health endpoints. This risk assessment provides an array of reasonable estimates for each endpoint based on the available epidemiological evidence. This array of results provides a limited degree of information on the variability and uncertainty in risk due to differences in study designs, model specification, and analysis years, amongst other differences. However, the second draft REA will provide a more comprehensive set of sensitivity analyses, especially for the short-term exposure mortality estimates, for which we only provide two sets of estimates based on the primary model specifications in the published studies.

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 REA that overstate the magnitude and certainty of ozone health risks. Comments on each of these issues are provided in the following sections.

1. Comments on options for the associations to model

The REA includes a mix of reported associations from multi-city studies as well as single-city studies. The criteria for selection include (1) the study is multicity and ideally, includes Bayes-adjusted city-specific effect estimates since these effect estimates combine local signals with broader regional or national signals, and (2) the study is not superseded by another study. There are two important cases where these criteria were not followed. Both Bell and Dominici (2008)[58] and Smith, Xu, and Switzer (2009)[59] provide regional analyses of ozone mortality associations that supersede the two main mortality analyses presented in the REA. In addition, the multi-continent APHENA study[60] 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

---

[58] M. Bell and F. Dominici, "Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 U.S. communities," *Am J Epidemiol* 167, 986-997 (2008).
[59] R. Smith, B. Xu, and P. Switzer, "Reassessing the relationship between ozone and short- term mortality in U.S. urban communities," *Inhal Toxicol*., 21, 37-61 (2009).
[60] K. Katsouyanni and J. Samet (2009*)*. "Air Pollution and Health: A European and North American Approach", (APHENA), *HEI Report* 142, Oct. 2009.

having a causal role in mortality or morbidity below the current standard. The authors of the REA were clearly aware of these studies because they are referenced in Chapter 7.

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 ISA and REA acknowledge that there is heterogeneity in ozone-mortality associations. However, the heterogeneity is much wider than EPA acknowledges and includes many cities with negative associations. The discussion of possible reasons for the heterogeneity in the ISA, REA, 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 6-27, 6-28, and 6-30 of the third draft ISA varies between -5 to +10% change in daily mortality for a 10 ppb increase in ozone.

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 REA. Toward that end, the second draft 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 must also be shown to policymakers. Figure 4 in Smith et al. (2009) demonstrates the differences for 8-hour ozone associations.

There is also strong evidence for unrecognized stochastic variability in associations within a given city. Ito (2003)[61] re-analyzed the 1220 separate air pollution mortality and morbidity associations that were included in the original Lippmann et al. (2000) HEI study of Detroit. As shown in Ito's Figure 2, there was a wide range of negative and 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.[62]

---

[61] K. Ito, "Revised Analyses of Time-Series Studies of Air Pollution and Health," *HEI Special Report*, pp. 143-156, May 5, 2003.

[62] F. Dominici et al., "Revised Analyses of Time-Series Studies of Air Pollution and Health," *HEI Special*

The same seasonal behavior is reported in the Medina-Ramon, et al. study of hospital admissions that is included in the REA, 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 REA 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 8, 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.

Report, pp. 9-24, May 5, 2003, at Figure 13.

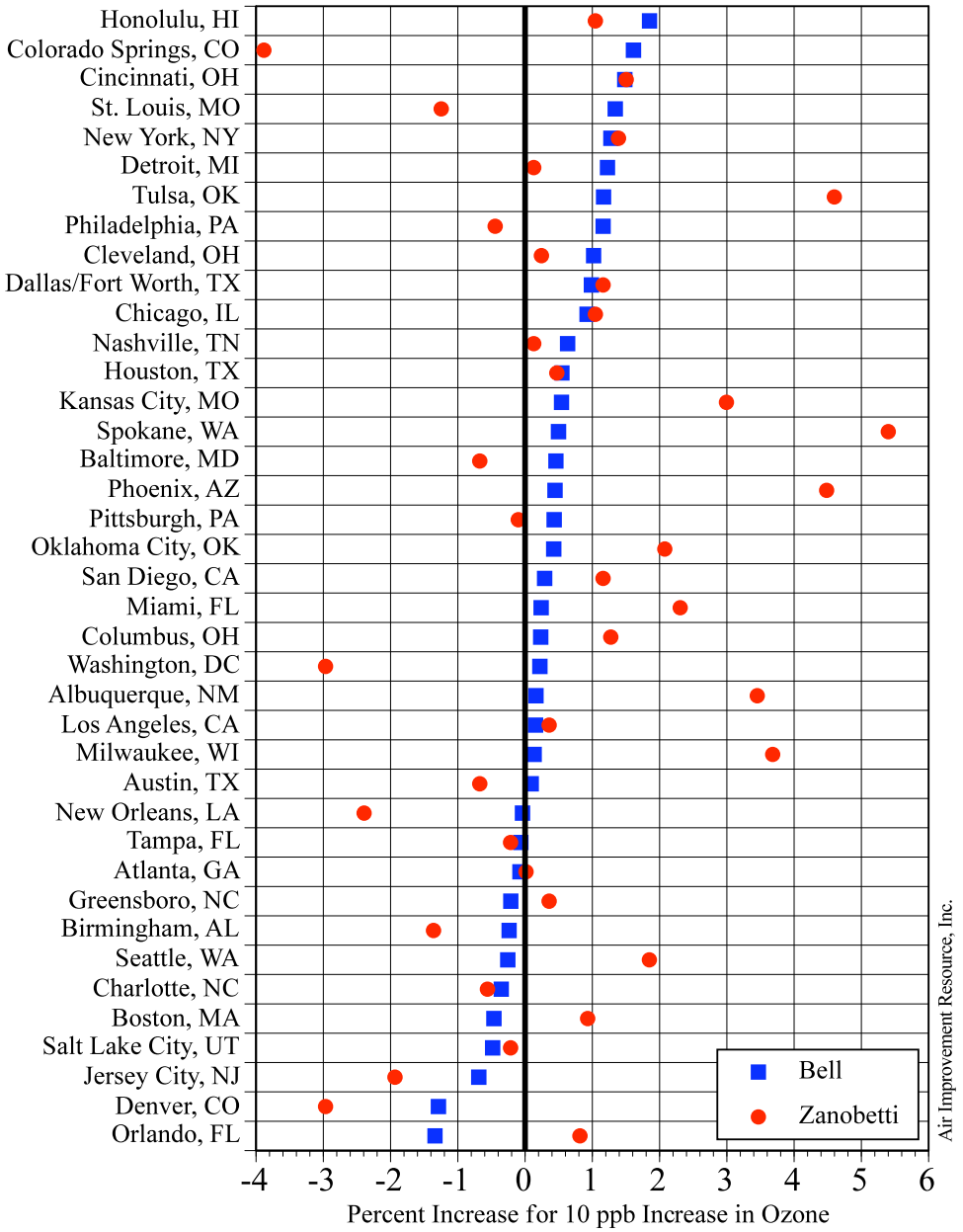


Figure 8 Comparison of unadjusted maximum likelihood estimates for mortality from Bell and Zanobetti.

Another demonstration of model uncertainty is given in Figure 9 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[63] 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 occurs before the peak ozone exposure. Again the wide variation in association for most cities is apparent in Figure 9.

[63] Ibid., at Figure 12.

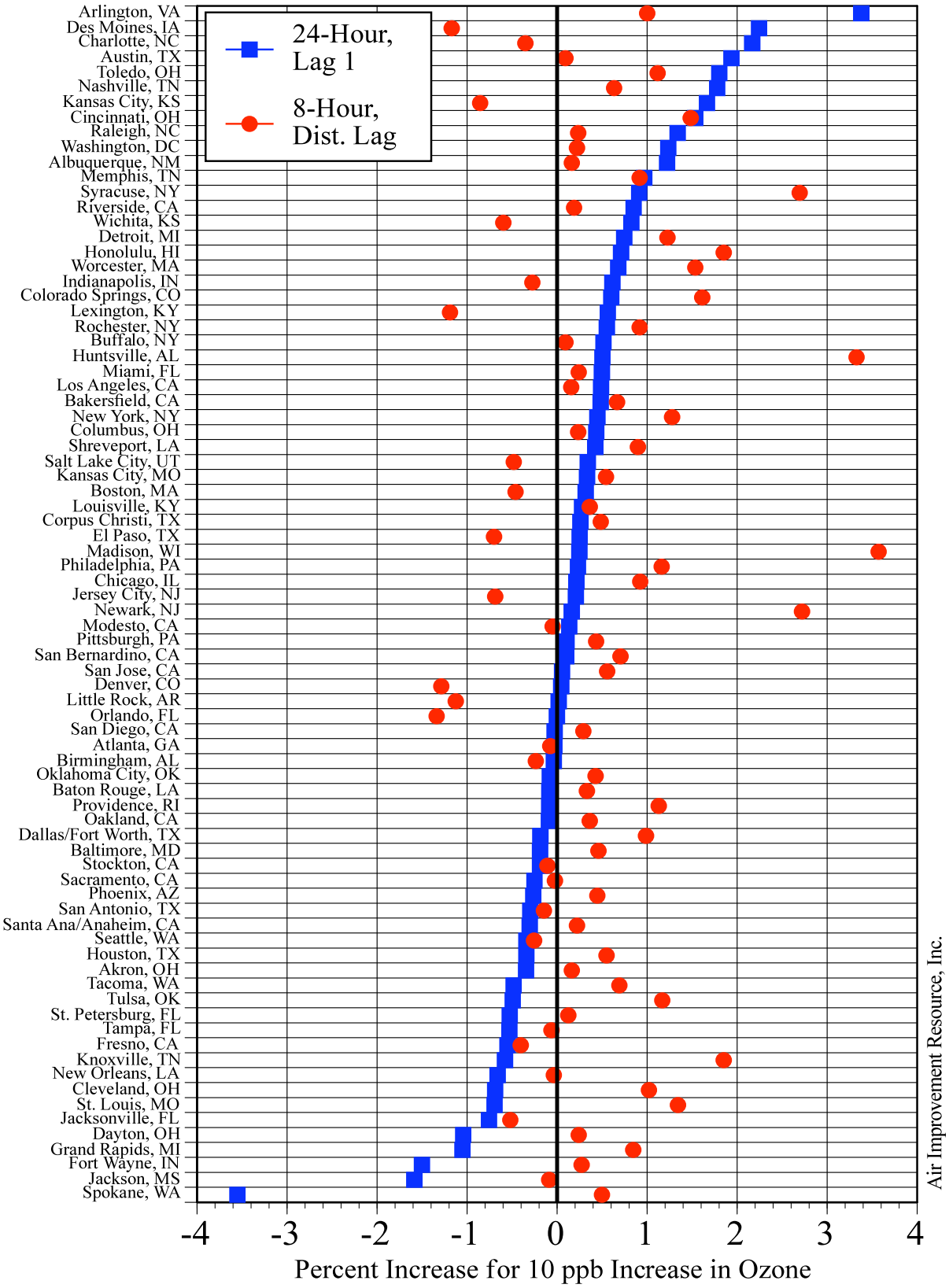


Figure 9 Maximum likelihood estimates for mortality from two NMMAPS analyses.

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:[64]

> 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 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 REA and PA and considered in the interpretation of the epidemiological data. Given that the small positive results from time-series studies may reflect residual bias of 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.

2. Comments on the shape of the C-R function

As discussed in the background ozone section, a key assumption that EPA makes in the REA is that there exists a no-threshold relationship between health effects and ozone concentrations down to a zero concentration. EPA's risk calculations are based on this assumption. As they explain in the ISA:

> The controlled human exposure and epidemiologic studies that examined the shape of the C-R curve and the potential presence of a threshold have indicated a generally linear C-R function with no indication of a threshold in analyses that have examined 8-h max and 24-h avg O_3 concentrations. However, there is less certainty in the shape of the C-R curve at the lower end of the distribution of O_3 concentrations due to the low density of data in this range.[65]

The last sentence, however, indicates there may be some uncertainty about this assumption at low ambient concentrations. In fact, they admit that sparse data at the low concentrations contribute to uncertainty when they say: "It is difficult to characterize the C-R relationship below 40 ppb due to uncertainty associated with the sparse data at these lower concentrations."[66]

---

[64] 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 Economicnagement*, 47, 30-54. (2004).
[65] U. S. Environmental Protection Agency, supra note 9, at p. 2-35.
[66] Ibid.

Further on, EPA presents a number of other reasons why a threshold may not be discernible in epidemiology studies:

> Generally, the epidemiologic studies that examined the O_3-mortality C-R relationship do not provide evidence for the existence of a threshold within the range of 24-h average (24-h avg) O_3 concentrations most commonly observed in the U.S. during the O_3 season (i.e., above 20 ppb). It should be noted that the evaluation of the C-R relationship for short-term exposure to O_3 and mortality is difficult due to the evidence from multicity studies indicating highly heterogeneous O_3-mortality associations across regions of the U.S. In addition, there are numerous issues that may influence the shape of the O_3-mortality C-R relationship that need to be taken into consideration including: multiday effects (distributed lags), and potential adaptation and mortality displacement (i.e., hastening of death by a short period). Additionally, given the effect modifiers identified in mortality analyses that are also expected to vary regionally (e.g., temperature, air conditioning prevalence), a national or combined analysis may not be appropriate to identify whether a threshold exists in the O_3-mortality C-R relationship.[67]

In addition, they acknowledge that exposure uncertainty, may even obscure thresholds in epidemiology studies:

> The various factors affecting exposure patterns and quantification of exposure result in uncertainty which may contribute to exposure measurement error in epidemiologic studies, which typically use fixed-site monitor data as an indicator of exposure. Low personal-ambient correlations are a source of exposure error for epidemiologic studies, tending to obscure the presence of potential thresholds, bias effect estimates toward the null, and widen confidence intervals, and this impact may be more pronounced among populations spending substantial time indoors.[68]

This issue is again revisited in Chapter 4 of the ISA:[69]

> Exposure misclassification can also tend to obscure the presence of potential thresholds for health effects, as demonstrated by a simulation study of nondifferential exposure misclassification (Brauer et al., 2002).[70]

and,

---

[67] Ibid, at p. 2-36.
[68] Ibid, at p. 2-14.
[69] Ibid, at p. 4-43.
[70] M. Brauer, J. Brumm, S. Vedal and A.J. Petkau, "Exposure misclassification and threshold concentrations in time series analyses of air pollution health effects," *Risk Anal.*, 22: 1183-1193 (2002).

> Nevertheless, low personal-ambient correlations are a source of exposure error for epidemiologic studies, tending to obscure the presence of potential thresholds, bias effect estimates toward the null, and widen confidence intervals, and this impact may be more pronounced among populations spending substantial time indoors.[71]

This view is consistent with points made by the Special Panel of the HEI Review Committee (Special Panel of the Health Review Committee, 2004)[72] 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). In addition, Rhomberg et al. (2011)[73] have recently shown, as others have previously shown, that measurement error can give a false linear result. Thus, the epidemiological studies cannot inform us as to whether there is or is not a biologic gradient for ambient ozone at low concentrations or whether there is or is not a threshold.

The toxicological studies that have been used to set chemical–specific standards demonstrate both threshold behavior and the presence of effects that not only become less common with progressively lower doses, but they also become less severe. A new study, which analyzed individual exposure-response data from 23 human controlled exposure studies for ozone that measured FEV responses under a wide variety of concentrations, activity levels and exposure patterns, found that the dose-response relationship was best described by a threshold model.[74,75] The existence of a substantial threshold for the first physiological effects in these controlled studies is not consistent with the assumption that the more severe effects suggested by some epidemiological studies have no threshold. Such assumptions are not consistent with either the general principles of toxicology or the specific findings of toxicological studies. Rhomberg et al. (2011) discusses these issues in detail:[76]

> The no-threshold proposal for noncancer toxicity is at variance with decades of experience in observing exposure-response relationships in pharmacology and toxicology, both within and below the usual

---

[71] U. S. Environmental Protection Agency, supra note 9, at p. 4-56.

[72] 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).

[73] L.R. Rhomberg, J. K. Chandalia, C. M. Long, and J. E. Goodman, "Measurement error in environmental epidemiology and the shape of exposure-response curves," *Critical Reviews in Toxicology,* 41:651-671 (2011).

[74] W.F. McDonnell, P.W. Stewart, M.V. Smith, C. Kim and E.S. Schelegle, "Prediction of lung function for populations exposed to a wide range of ozone conditions," *Inhalation Toxicology,* 24:619-633 (2012).

[75] E.S. Schelege, W. C. Adams, W.F. Walby and M.S. Marion, "Modelling of individual ozone exposure response kinetics," *Inhalation Toxicology,* 24:401-415 (2012).

[76] L.R. Rhomberg, J. Goodman, L.Haber, M. Dourson, M. Andersen, J. Klaunig, B. Meek, P. Price, R. McClellan and S. Cohen, " Linear low-dose extrapolation for noncancer health effects is the exception, not the rule," *Crit.Rev.Toxicol*., 41:1-19 (2011).

> experimental range for environmental chemicals.

They further note:

> The no-threshold idea is also belied by our experience with medicines, poisons, foodstuffs, and many other kinds of exposure to agents that can have toxic effects if experienced in excess. With the possible exception of allergic reactions, within the range of low exposures, we do not observe slightly increased exposures to such agents somewhat increasing the probability that we will suffer the full effect of a toxic dose. In therapeutics, a small fraction of the therapeutic dose will not necessarily produce a moderate or full response in a diminished fraction of the treated population. It is only when the critical concentration is sustained at the site of action for the necessary period of time that an effect will be elicited. The experience of exposure thresholds for biological effects, including adverse effects, pervades daily life.

They also argue that the no-threshold proposal is at variance with basic tenets of homeostasis—the robust nature of living systems.

In summary, the shape of the concentration-response is not known and epidemiology studies cannot be used to identify threshold because of exposure uncertainty. Consequently EPA's extrapolations of risk at low ozone concentrations in the REA are not justified.

3. Comments on the data presentation in the REA

The risks estimated in the REA are presented in a series of tables in Chapter 7 and in maps in Chapter 8. The risks are typically presented as point estimates with the confidence levels included in some of the tables. In many key cases, the confidence limits include zero. For example, the lower confidence bounds for ozone mortality at current ozone levels for 8 of the 12 cities in Tables 7-11 and 7-12 are negative. Similarly, the lower confidence bounds are negative for 10 of 12 cities in Tables 7-13 and 7-14. The REA indicates:

> Population incidence estimates with negative lower-confidence bounds do not imply that additional exposure to O_3 has a beneficial effect, but only that the estimated O_3 effect estimate in the C-R function was not statistically significantly different from zero.

Given the lower confidence bounds in so many cities are negative, the REA should highlight that for most of the 12 cities we cannot conclude that there is a true ozone mortality association much less an ozone mortality health effect. Instead, the REA presents the point estimates from the 12 cities in additional tables and makes an argument that negative values for lower bound statistics do not imply that O_3 is beneficial, but rather speak to too low a sample size. This argument demonstrates a major bias in the REA and PA – the authors start with the assumption that ozone is causing mortality rather than evaluating and weighing the full range of evidence.

D. Estimates of variability and uncertainty in the risk assessment

The fourth and fifth goals are to understand the influence of various inputs and assumptions on the risk estimates and the uncertainties in those estimates. This is laudable and particularly important.

For the clinical studies, the REA indicates that the most influential elements of uncertainty in the exposure assessment are the activity patterns, air exchange rates, the spatial variability in ozone concentrations, the metabolic equivalents of work distributions, and the resting metabolic rate and ventilation rate equations. The draft indicates that the second draft REA will include the results of sensitivity analyses for each of these five elements.

One key issue will be the activity patterns and predicted ventilation rates used for various cohorts, such as asthmatics and outdoor workers. AIR urges the Agency to compare the model estimates with the data in the literature. For example, Shamoo et al., 1991[77] investigated the summer activity patterns of outdoor workers in Los Angeles and 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. In a related study, Shamoo et al., 1994[78] investigated the time-activity patterns of asthmatics in the Los Angeles area. In the case of medium activity (mean ventilation rate of 37 L/min for men and 24 L/min for women), outdoor medium activity accounted for 19 minutes per day on average. Outdoor fast activity (mean ventilation rate of 61 L/min for men and 32 L/min for women) occurred 2 minutes per day on average. There are also studies of activity patterns and ventilation rates of elementary and high school students in Los Angeles during the ozone pollution season[79] and data on inhalation rates of ruse in risk assessment for construction workers that can be used to provide reality tests on the APEX VR and EVR output.[80]

With regard to the spatial variability of ozone, AIR urges EPA to include a sensitivity case for the vertical variation in ozone between measurement height and person height. As noted above, there is concern that the ventilation rate algorithms predict too many high ventilation rates in APEX and there is a separate concern with the binning of all EVR rates of 13 and above. Both these issues should be thoroughly evaluated in the final REA.

The discussion of uncertainty in Chapter 6 indicates that one of the most important sources of uncertainty is the population distribution of estimated 8-hour daily maximum O_3 exposure concentrations in each urban study area. Chapter 6 notes that the uncertainty regarding these

---

[77] 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).
[78] D. Shamoo, W. Linn, R. Peng, J. Solomon, T. Webb, J. Hackney, and H. Gong, Time-activity patterns and diurnal variation of respiratory status in a panel of asthmatics: implications for short-term air pollution effects," *J. Exp. Anal. Environ. Epidemiol*., 4, 133-148 (1994).
[79] C. Spier, D. Little, S. Trim, T. Johnson, W. Linn, and J. Hackney, Activity patterns in elementary and high school students exposed to oxidant pollution, *J. Exp. Anal. Environ. Epidemiol*., 2, 277-293 (1992).
[80] M. Allan, H. Jones-Otazo, and G. Richardson, "Inhalation rates for risk assessments involving construction workers in Canada," *Human and Ecological Risk Assessment*, 15, 371-387 (2009).

estimated exposures is discussed in Chapter 5 and that they are not discussed further in Chapter 6 and, importantly, that their uncertainty is not propagated through the risk calculation. Therefore, in this draft, only uncertainties in the exposure-response functions are considered for inclusion in the second draft REA.

This is a major mistake and omission. Since the overall risk depends on the exposures calculated by APEX, omitting their uncertainty would lead to an overly precise estimate of risk based on the clinical studies.

In addition to the issues for the clinical studies, the REA should evaluate the wide range of positive and negative associations in individual cities and the temporal and spatial variations in the combined associations in multi-city studies. For example, in the regional analysis by Bell and Dominici, 2008 only two of seven regions have positive and statistically significant ozone/mortality associations while two have small negative associations. 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 0.030 ppm increase in 8-hour ozone. The individual-city associations for pneumonia hospital admissions ranged from –15% to + 20% for a 0.030 ppm 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 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. By exploring the full range of spatial and temporal differences in association together with model selection uncertainty, the limitations of the epidemiologic risk assessment will become apparent.

E. Estimates of the national mortality burden

The sixth goal is to understand the national mortality burden associated with ozone. Chapter 8 presents several maps that plot the results from applying the Bayesian-adjusted city-specific mortality estimates together with the national Bayesian-adjusted estimates. For example, Figure 8-5 presents the point estimates of non-accidental deaths attributable to ozone for each county. Figure 8-7 present the results as a percent of total mortality for each county. The Figures give the impression that people are dying in every county around the country and that there is a large swath of the Western U. S. with the highest death toll, between 2 and 2.4 % of total mortality. AIR is concerned that these figures give a highly misleading impression of the risk from ozone when the variability and uncertainty in the underlying data are taken into account.

If, rather than use the Bayesian combined estimates from Bell et al. (2004), a variety of other possible approaches and baselines are considered, a very different picture emerges. For example, if the regional combined results from Bell and Dominici (2008) are considered, the ozone mortality impact varies dramatically across the country, with most of the country having no statistically significant ozone mortality association. In fact in the regional analysis, some of the portions of the country that look the worst in the REA maps, actually have negative combined associations. This is shown in Figure 10. Smith et al. (2009) present similar maps for their regional analyses, with any ozone mortality effect confined to the Northeast section of the country.

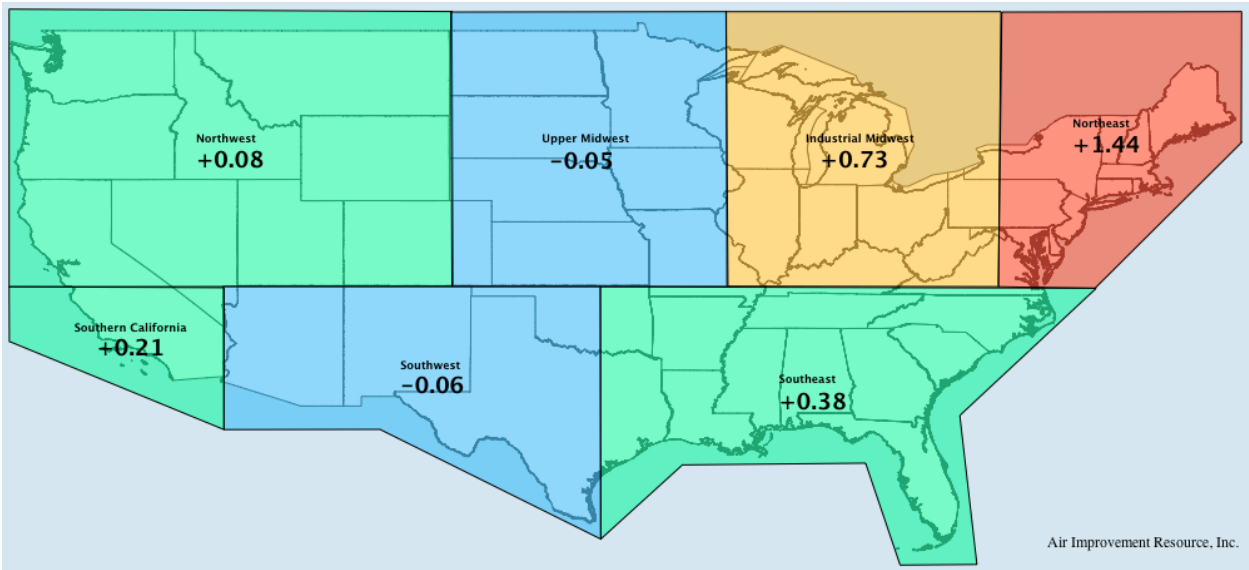


Figure 10 Combined associations (percentage increase in mortality for a 10 ppb increase in ozone) in 7 regions from Bell and Dominici (2008); blue areas are negative and non-significant, green areas are positive and non-significant, yellow and red areas are positive and significant.

If the individual city results from Bell et al., 2004 are considered, the pattern includes both positive and negative cities, a finding which is not biologically plausible. This is demonstrated in Figure 11 which show a patchwork of cities that have positive and negative ozone mortality associations in each of the seven regions used by Bell and Dominici as well as Smith et al.

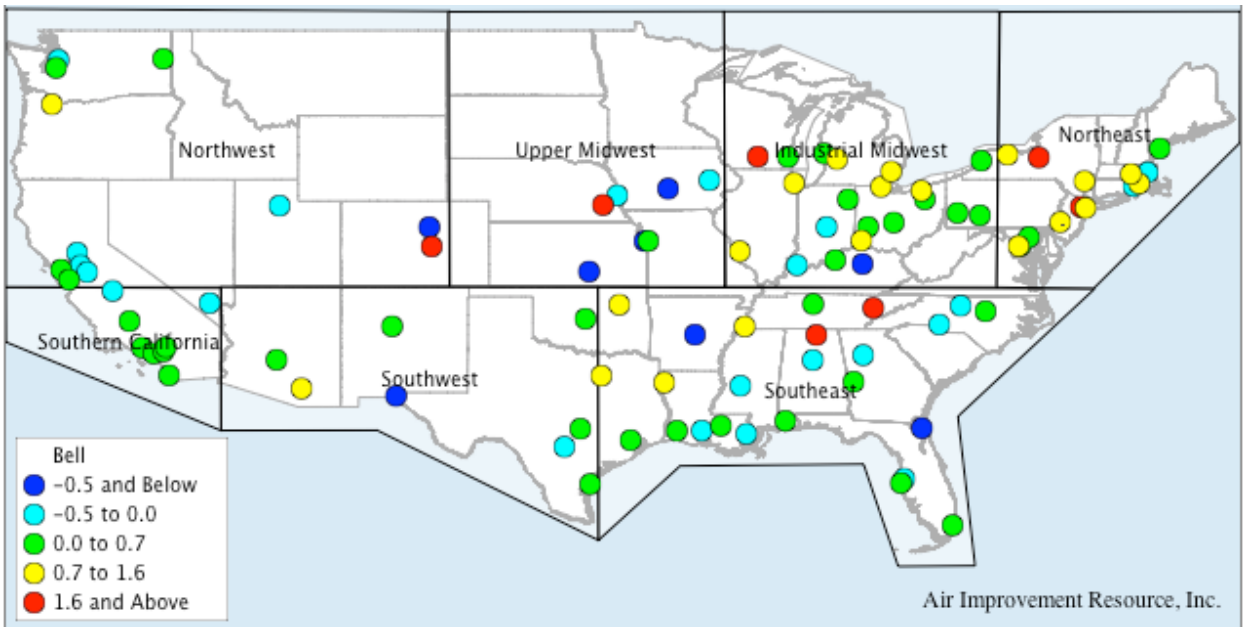


Figure 11 Unadjusted maximum likelihood estimates of mortality from Bell et al. (2004)

Since the REA includes individual city associations for morbidity, it should also include the

unadjusted individual city results from Bell et al. (2004) shown in Figure 11 for mortality. If these associations are substituted for the Bayesian-adjusted data, the conclusion would be drawn that there are counties where ozone is associated with a wide range of mortality outcomes, from increasing mortality by up to 14 % in some counties to decreasing total mortality by up to 6 % in other counties. This range is not biologically possible and should cause EPA staff concern with any estimates of a nationwide mortality burden from ozone exposure.

AIR would encourage EPA to evaluate two other potential baseline estimates. One would be the NMMAPS data presented in the 2003 HEI Special Report. We suggest obtaining the winter NMMAPS data referred to above that shows a negative combined association, and the summer NMMAPS data from multi-pollutant models as shown in Figure 12 of the Dominici material in the Special Report. None of the summer ozone associations in multi-pollutant models were statistically significant for lag 0, 1 or 2. Based on these data, the conclusion would be that we cannot say whether ozone has an independent effect on mortality.

The second recommendation is to use the combined results from the APHENA study for all-cause mortality with PM controlled. When this analysis is considered, none of the 16 model combinations had a positive and statistically significant association in the United States.[81]

With all these possible choices for the analysis, the futility of estimating a national mortality burden becomes apparent. Since the ISA does not conclude that the relationship of ozone with all-cause mortality is causal, EPA should not include claims that ozone causes 18,000 deaths a year as presently given in Table 8-2.

F. Comments on Chapter 9

Chapter 9 is designated as a synthesis, but it is not a synthesis in its current form. The first sections are simply summaries of the various risk analyses in the body of the REA. The last section lists some observations about the analyses and factors that may change with the second draft.

There is one overarching issue when one compares the risks estimated from clinical studies and epidemiological studies that is critical for the NAAQS review but not mentioned in the synthesis chapter. There is a major disconnect between the results of the controlled studies which we know are causal and the results implied by the observational studies that EPA relies on. The existence of a substantial threshold for the first physiological effects in controlled studies is not consistent with EPA's assumption that the dramatically more severe effects suggested by some epidemiological studies have no threshold. Such an assumption is not consistent with either the general principles of toxicology or the specific findings of ozone toxicological studies. The risk at the current standard as estimated from the controlled exposures is minimal. However, the REA makes the claim that levels below the current standard are causing substantial mortality. The disconnect between these two estimates of risk should be acknowledged in the REA and discussed fully in the PA.

[81] AIR comments, supra note 2, at pp. 25-26.

III. Comments on the PA

A. Comments concerning process

Although, in the normal course of events, the first draft PA would be prepared after the ISA is complete and after the REA has had CASAC and public review, the draft PA was prepared before the ISA was finished and before any external review of the first draft REA. This is an extraordinary circumstance. Since the PA refers to what are incomplete documents, any changes in those documents as they become final would need to be carried over to the next draft or the final PA. As documented in Sections I and II of these comments, major revisions are necessary in the REA. Thus, major revisions will also be called for in the relevant portions of the PA.

The revisions that are necessary in the REA will have a major effect 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. Thus, it is premature in the PA to draw any conclusions as to the adequacy of the current ozone standard.

Since the draft PA was written without any CASAC or public input on the REA or the final CASAC and public input on the ISA, it represents solely EPA staff thinking. 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.

B. Comments concerning content

1. Comments on Chapter 2 - Consideration of the health evidence

Chapter 2 summarizes a great deal of information concerning health effects from ozone. Based on AIR's review, Chapter 2, in a number of key locations, overstates the consistency and coherence of the evidence. One of the most important examples is the question of ozone and respiratory mortality. The PA includes statements such as:[82]

> Recent evidence from several multicity studies and a multicontinent study also demonstrate consistent positive associations between short-term exposure to ambient O_3 concentrations and increases in respiratory mortality.

Appendix 1 to these comments demonstrates, in contrast, that 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 addition, the Health Effects Institute provided comments noting that the ISA oversimplifies the APHENA findings, focuses on selected results, and draws stronger conclusions than would the investigators or the HEI Review Committee.[83] The HEI comments pointed out that this is especially true of the analyses of

[82] PA, supra note 2, at pp. 2-74.
[83] December 29, 2011 letter from D. Greenbaum to Drs Vandenberg and Samet, Health Effects Institute Comments on Docket EPA-HQ-ORD-2011-0050, Second Draft Integrated Science Assessment on Ozone, at page 2.

respiratory vs. cardiovascular mortality and the lack of coherence between the mortality and hospitalization analyses. After providing detailed comments supporting these general comments, HEI indicated:[84]

> We would suggest that, given that the APHENA study is the single major multi-city analysis of air pollution and mortality published since the last ISA, that NCEA carefully review – as described above – its treatment of APHENA, and especially the degree to which the relative lack of coherency between the mortality and hospitalization results affects the conclusions that NCEA can draw on causality.

The HEI Review Committee concluded, in the APHENA report, that in all-year analyses associations between ozone and respiratory mortality were generally close to zero and not significant in any region or in the combined estimate for all three regions. The APHENA investigators themselves concluded that there was little evidence for an effect of ozone on respiratory mortality in any center. As noted in Appendix 1, while associations were generally higher in summer-only analyses in the U. S., only 2 of 12 model combinations were statistically significant and, when controlled for PM10, none of the 8 model combinations presented in the APHENA report were statistically significant.

Despite the detailed criticism in various public comments and the different interpretation of the APHENA results by the investigators and the HEI Review Committee, EPA continues to overstate the magnitude, consistency, and coherence of the APHENA findings.

In another example in which the PA overstates the consistency of epidemiological findings, the PA concludes:[85]

> Recent epidemiologic time-series studies that include additional multicity studies and a multicontinent study further support that short-term exposures to ambient O_3 concentrations are consistently associated with increases in respiratory hospital admissions and emergency department visits specifically during the warm/summer months in multiple geographic locations and across a range of O_3 concentrations.

In contrast, the various multi-city studies all show that the individual city associations range from positive to negative, which is biologically implausible. In addition, in the largest study, APHENA, none of the four models assessing respiratory hospital admissions in the summer in the U. S. were statistically significant.

Based on the overstatements of consistency and coherence, the draft PA concludes that there is a causal relationship between short-term O_3 exposure and a full range of respiratory morbidity effects, including hospital admissions and ED visits, which provides support for concluding that short-term O_3 exposure is associated with incapacitating effects. Further, the PA states:[86]

---

[84] Ibid., at pp 6.
[85] PA, supra note 2, at pp 2-74.
[86] Ibid., at pp 2-51.

> Overall, the evidence supporting an association between short-term O_3 exposures and respiratory mortality is much stronger.

The concluding section of Chapter 5 first asks then answers the question:[87]

> To what extent has scientific information become available that alters or substantiates our understanding of the health effects that occur following short-term or long-term exposures to O_3, and our understanding of the O_3 concentrations at which such effects occur?

Although this is the most important question for the PA to answer, the answer in the draft PA is biased, as discussed above, to overstate the strength and consistency of the findings since the previous review was completed. Instead, the PA should note a number of new findings that must influence the interpretation of the human clinical and epidemiological data.

First, there are now several new clinical studies of the first physiological effects of ozone at low concentrations with exercise. As documented in Section II of these comments, the effects reported at 0.06 ppm are small and not adverse according to the American Thoracic Society guidelines for adversity. The experimental protocols in these studies involve 8-hour ventilation rates that are at the high end of those experienced in the real world. There are now two analyses of the FEV1 data that conclude that there is a threshold for these first effects. Thus, instead of the previous EPA assumption that these effects are linear down to zero, the PA should consider the new McDonnell et al. (2012) concentration-response function model.

Second, the existence of a substantial threshold for the first physiological effects in controlled studies is not consistent with EPA's assumption that the dramatically more severe effects suggested by some epidemiological studies have no threshold. Such an assumption is not consistent with either the general principles of toxicology or the specific findings of ozone toxicological studies. The PA should address the issue of dose plausibility in detail.

Third, there are now several major multi-city studies of hospital admissions and mortality that all demonstrate a pattern of results that is not consistent with ozone causality. The unadjusted individual city results in these studies cover a biologically impossible wide range from positive to negative. The combined associations tend to be slightly positive in the summer but slightly negative in the winter. The largest of these studies, APHENA, raises issues with the EPA claims of coherence and consistency in the observational data. The ISA and draft REA ignore the negative associations in the literature, the proverbial elephant in the room, by looking only at the combined associations in multi-city studies and emphasizing EPA-favored positive associations in single-city studies. EPA favors combined associations when it likes the result, but raises concerns when it wants to downplay a study's results. For example, in discussing the Stylianou and Nicolich (2009) study that reported thresholds in ozone mortality relationships, the PA notes that "given the city-to-city variation in risk estimates, combining the city-specific estimates into an overall estimate complicated the interpretation of the results."[88] In addition, the issues of model selection uncertainty, confounding, and publication bias are ignored or downplayed in the

[87] Ibid., at pp 2-72.
[88] Ibid., at pp 2-34.

various EPA documents. As Figures 8 and 9 in Section II demonstrate, the uncertainty due to model selection issues is extremely large. The PA should address all these issues in the interpretation of the observational studies and integration of the full range of ozone effects studies.

Fifth, as documented in Section I, the understanding of the role of background ozone, that due to non-U. S. anthropogenic sources, has changed substantially since the prior review. Estimates of both the means and extremes of background are now substantially higher. The EPA should revise the consideration of background in the PA as discussed in Section I.

2. Comments on Chapter 3 - Assessment of ozone exposures and risks

Chapter 3 in the draft PA is simply a summary of the key findings in the draft REA. As documented in Section II, by omitting the estimated exposure results for person-days, the draft REA gives a misleading impression of the risks based on the clinical studies. In addition, by omitting consideration of the full pattern of associations in the epidemiological studies and omitting consideration of model selection uncertainty, the REA and PA overstates the magnitude, consistency, and coherence of the epidemiological evidence for ozone health effects.

When the sensitivity analyses recommended in Section II are completed, the REA conclusions and key findings will change substantially. Thus, Chapter 3 in the second draft PA will, of necessity, change substantially.

3. Comments on Chapter 4 –Preliminary staff conclusions on the adequacy of the current standard

Based on the draft ISA and the draft REA, the draft PA discusses the available evidence and draws several conclusions. The first is that "the available evidence provides strong support for a standard at least as protective as the current O_3 standard."[89] The second is:[90] that:

> In addition, we reach the further preliminary conclusion that the available evidence calls into question the adequacy of the current standard and provides support for considering potential alternative standards to increase public health protection against the effects related to short-term O_3 exposures, especially for at-risk groups.

AIR is concerned that EPA has reached this preliminary conclusion before receiving any public or CASAC input on the REA and before the ISA has been finalized. As documented in these comments, the draft REA and draft PA overstate the case for ozone health effects.

In particular, 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 for 8-hours at a rate that is at the very high end of real-world situations. Based on the APEX estimates of the number of person-days of exposure above EPA's benchmarks with an even

[89] Ibid., at pp 4-29.
[90] Ibid.

lower level of exercise, the fraction of person-days experiencing such effects is extremely low. Thus, these are rare occurrences at current ozone levels and even rarer occurrences when the current standard is attained.

In reaching the preliminary conclusion that the current standard may be inadequate, the PA indicates:[91]

> In some individuals, the types of O_3-induced respiratory responses reported in controlled human exposure studies could become severe enough that they result in increased use of medication, emergency room visits, and/or hospital admissions. Thus, the strong evidence for lung function decrements, respiratory symptoms, airway inflammation, and other respiratory effects following exposures to O_3 concentrations commonly encountered in U.S. urban locations supports the biological plausibility of the conclusions that exposures to ambient O_3 concentrations can result in respiratory-related hospital admissions and emergency department visits, as well as the most severe O_3-associated effect, premature mortality.

This argument is purely speculative as it relates to ozone concentrations at or below the current standard. There is no evidence that subjects – normal, asthmatic, or otherwise respiratory-compromised -- have experienced any of these complications even after exposures to ozone in human clinical studies that are substantially higher than the current standard.[92]

The PA discussion regarding adequacy also places considerable weight on the evidence for mortality effects.[93] In addition, the PA re-states the CASAC advice regarding the level of the standard from the last review and notes:[94]

> Since this advice was provided, based on evidence available in the last review, the evidence for adverse health effects following short-term exposures to O_3 concentrations below 75 ppb has become even stronger, with the addition of several controlled human exposure and epidemiologic studies conducted at relatively low O_3 concentrations. Given this, we note that, at a minimum, nothing in the recent evidence would contradict CASAC's previous advice and that, in fact, recent evidence provides stronger support for that advice.
>
> In light of all of the above considerations, staff reaches the preliminary conclusion that the body of information now available supports consideration of revising the current 8-hour O_3 primary standard, so as to afford greater public health protection against the adverse health effects of short-term O_3 exposures, especially to at-risk groups, and that it does not support retention of the current standard.

AIR submits that the evidence for adverse health effects from ozone has not become stronger

---

[91] Ibid., at pp 4-14.
[92] I. Mudway and F. Kelly, An investigation of inhaled ozone dose and the magnitude of airway inflammation in healthy adults, *Am J Respir Crit Care Med*, 169: 1089-1095 (2004).
[93] PA, supra note 2, at pp. 4-29.
[94] Ibid., at pp 4-45.

since the last review. During the last review, CASAC also advised the Administrator 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."[95] CASAC was also concerned that the degree of ozone measurement error would be expected to have a substantial impact on the ability to detect a threshold in the concentration-response relationship, noting "pollutant exposure measurement error obscures true thresholds in the concentration-response relationship, and this effect worsens with increasing degrees of measurement error."[96]

AIR is concerned that EPA is relying on CASAC's previous advice regarding the level of the standard and is not considering the new information that (1) background ozone is much closer to the current standard than thought during the last review, (2) we now have 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.

All of these factors may indeed contradict CASAC's previous advice regarding the level of the standard. Therefore, it is premature to draw any conclusions regarding the adequacy of the current standard until the REA is completed.

[95] R. Henderson, CASAC Letter, EPA-CASAC-06-07, June 5, 2006 at pp. 3.
[96] Ibid., at pp. 4.

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)*[97] 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 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

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

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 ≥ 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.

Appendix 2 - Issues Encountered with the APEX Code

According to the documentation, the APEX code conforms to the latest FORTRAN standards. However, this is not entirely true. In trying to compile the code using the latest (2013) version of the Absoft compiler, AIR encountered the following issues:

1) Use of intrinsic functions with mismatched argument types. For example consider: X=MIN(Y,A), where X and Y are double-precision variables (REAL*8), but A is single-precision (REAL*4). The FORTRAN standard requires that both arguments must be of the same type and precision.
2) Use of integer intrinsic functions in logical expressions. For example consider: IF (INDEX(STRING,'ABC')) A=B. The intrinsic function INDEX returns an integer value. The FORTRAN standard requires that LOGICAL results be used in this IF context.
3) Use of mixed variable types and precision in calculations. For example consider: Z = 13/N + A/Y – EXP(1/B), where N is an INTEGER, A and B and the intrinsic function EXP() are single-precision (REAL*4), and Z and Y are double-precision (REAL*8). Although most compilers will automatically convert the expression components so that all calculations are performed using double-precision, this conversion is not part of the FORTRAN standard.

Since the compiler flagged these Issues 1) and 2) as errors, the code had to be corrected before it could be used. The Absoft compiler automatically promotes expressions to the highest order, so Issue 3 was not flagged. Based on this experience, other uses may run into similar problems using the code.