

**Comments on the  
First Draft OAQPS Staff Paper  
“Review of the National Ambient Air Quality Standards for Ozone:  
Policy Assessment of Scientific and Technical Information”  
EPA-452/D-05-002, November 2005**

**Prepared for  
Alliance of Automobile Manufacturers**

**by**

**Jon M. Heuss  
Air Improvement Resource, Inc.  
January 16, 2006**

## **Executive Summary**

Based on a review of the draft Staff Paper (SP), AIR has identified the following key concerns.

- EPA has underestimated policy relevant background, the ozone that is not subject to control by reduction of man-made U. S. emissions. The model EPA relies on is not capable of resolving the phenomena that determine attainment or non-attainment of an extreme value standard throughout the country. A standard set as low as 0.06 ppm would put virtually the entire country out of attainment.
- EPA has overestimated the magnitude and consistency and underestimated the uncertainty in the results of acute epidemiologic studies, especially for school absences, hospital admissions, and mortality. EPA has also overstated the biologic plausibility of the epidemiologic associations for serious health outcomes. While there is now a much larger set of studies than was available in 1996/7, there is an implausibly wide range of results from positive to negative in systematic analyses, and issues of model selection and publication bias cloud the interpretation of the data.
- The acute respiratory effects in controlled human studies identified during the 1996/97 review are still relevant and offer the best information on ozone effects. However, the risk from these effects must be evaluated taking into account the distribution of personal exposures and exertion levels in the population.
- There are no substantive chronic effects of ozone in humans that have been identified. This is most likely a result of the low mean personal exposures of the population even in areas that substantially exceeded the 1-hour and 8-hour ozone standards in the past.

- There is no new data on vegetation effects suitable for use in establishing secondary standards. The planned risk assessment will utilize the crop yield loss and seedling growth data used in the previous review. Although more-biologically based flux-based models offer promise, there is insufficient data to apply them at present.
- It is not clear if the planned vegetation risk assessments will reduce the major uncertainty that was acknowledged in the 1996/1997 review. There is still a paucity of ozone monitoring in crop growing regions, the model EPA plans to use to aid in assessing rural exposures is itself subject to major limitations and uncertainties, the uncertainties in extrapolating from effects in open-top chambers to effects in natural settings are still a concern, and the limitations and simplifications involved with cumulative ozone metrics are still a concern.

Based on our review and comments, the second draft SP should be extensively revised to correct the errors, omissions, and overstatements we have identified.

In Chapter 2, the SP should:

- Include a range of estimates for background ozone (both mean and peak levels) and acknowledge the limitations of the one modeling study used in the first draft SP to provide background ozone for the risk estimates
- Include maps similar to the maps we have included in the body of our comments to enable the reader to evaluate the spatial distribution of various ozone metrics relevant to evaluating the health and welfare studies as well as alternative ozone standards.

In Chapter 3, the SP should:

- Remove statements by which the SP overstates the magnitude and certainty with which the health effects of ozone have been established in epidemiological studies. As documented below, the SP overestimates the magnitude and consistency and underestimates the uncertainty in the results of acute epidemiologic studies, especially for symptoms, school absences, hospital admissions, and mortality.
- Add explicit recognition that, even though there is a larger epidemiologic database than in 1996, in systematic analyses there is an extremely wide range of results in individual cities ranging from strongly positive to strongly negative.
- Add explicit discussion indicating that publication bias is a major concern leading to overestimates of the strength and consistency of the epidemiologic database on ozone effects.
- Revise the discussion of model selection to indicate that, relying on the confidence limits from a single regression model severely underestimates the true uncertainty in the statistical evidence.

- Add explicit discussion of the role of the body's defense mechanisms against an irritating gas such as ozone as they relate to the medical significance of single or repeated exposures to ozone at various levels and the plausibility of various health effects implicated by epidemiologic studies. These defenses include the presence of antioxidants in the epithelial lining fluid, a neurally-mediated reflex action to inhibit maximum inspiration, and initiation of immune system responses that resolve within a few days and at higher doses tissue damage and repair mechanisms that resolve over longer time periods. The controlled studies in the CD clearly demonstrate that the mechanisms through which ozone exerts toxic effects are non-linear. The fact that lesser effects are observed earlier and at lower exposures than more severe health effects is an important consideration when it comes to evaluating the plausibility of the epidemiologic associations reported in the literature.

In Chapter 4, the SP should:

- Evaluate the APEX model's performance versus ozone personal exposure data. The Agency should also evaluate the model's intermediate outputs for key variables that influence the final result. These include the distribution of time outside as well as the distribution of equivalent ventilation rates (EVR). It is particularly important to evaluate how the model's predictions of the number of occurrences of elevated ventilation rates compare with real world data.
- Develop dose-response curves as a function of EVR to avoid the over-estimation of risk. There is evidence that both the dose-response curve and the distribution of exposure varies with EVR. We expect many more 8-hour occurrences of ozone exposures at 13 EVR than at 20 and many more at 20 than at 27. Thus, the approximation that all exposures between 13 and 27 EVR respond as though they were at 20 EVR will substantially overestimate the risk.

In Chapter 5, the SP should:

- Re-visit the shape of the dose-response curve for lung function decrements allowing for non-linear responses and including new data from low concentration experiments (below 0.08 ppm). The use of linear dose-response is problematic since evaluations of the FEV and other acute respiratory responses have been found to be non-linear. The use of a linear response will overestimate the risk based, not on data of effects, but based on an unsupportable assumption. The full details of the methodology need to be made available to both CASAC and the public.
- Include explicit consideration in the risk assessment for the epidemiologic associations of a range of assumptions concerning causality, background levels, and shape of the dose-response. In addition, the full range of associations in systematic analyses from positive to negative should be included to demonstrate the uncertainty in the epidemiologic data base and the implications of both protective and harmful effects of ozone that vary with study, city and season.

In Chapter 6, the SP should:

- Concentrate on the respiratory effects of acute ozone exposures as determined in controlled experiments. The risk from these exposures should be evaluated with a validated exposure model and a revised lung function decrement dose-response curve.
- Discount the epidemiological evidence. The large epidemiological database should be discounted because (1) there is a biologically implausible wide range of results (from positive to negative) in systematic analyses, (2) the uncertainty due to model selection is substantially larger than reported in single pollutant models, (3) publication bias leads to an erroneous impression of the consistency and strength of the database, and (4) all pollutants show similar associations for hospital admissions and mortality.

In Chapter 7, the SP should:

- Explicitly consider and discuss the major sources of uncertainty in the vegetation risk assessment identified during the 1997 rulemaking. Since the significant uncertainties in the previous review led the Administrator to conclude that the benefits of alternative secondary standards should be regarded as rough approximations, the quantitative treatment and evaluation of the uncertainties will be particularly important.
- Include a more extensive analysis of the capability of CMAQ or other possible models to simulate rural ozone, in general, and any seasonal ozone metrics under consideration. Existing data for CMAQ and other regional models raise serious issues including an inability to simulate observed elevated ozone events with many false positives and false negatives, poor performance for ozone precursors and nitrogen dioxide photolysis rate, and uncertainty in vertical mixing and dry deposition rates.
- Include specific analyses of the sensitivity of crop loss estimates to various assumptions concerning background and to a correction for differences between ozone at measurement height and ozone at plant height.

## **Introduction**

Air Improvement Resource, Inc. (AIR) reviewed the draft Staff Paper (SP) with a focus on the characterization of scientific information that will affect the choice of appropriate ozone air quality standards. The main topics of interest are the policy relevant background of ozone, the interpretation of epidemiologic studies of ozone, the integrated synthesis of information from controlled human studies, toxicology, and epidemiology, and the risk assessments for human health and vegetation effects. In the following sections, we provide comments on each of these topics. We provide both general and detailed comments on portions of the text of the draft SP that do not accurately reflect the scientific information relevant to setting ozone air quality standards.

AIR provided comments on the first and second draft Criteria Documents (CD).<sup>1</sup> Since the second draft CD was undergoing review by the Clean Air Scientific Advisory Committee (CASAC) and the public when the draft SP was issued, the references to material from the second draft CD in the SP are subject to change. When the final CD is issued, staff will have to review and revise the relevant sections of the SP. Based on public and CASAC comments provided to the Agency at the December 6-7, 2005 CASAC meeting, the interpretation of the epidemiologic studies in the CD needs to be substantially revised pointing out the major weaknesses and problems with this database. After that revision, as noted above, the draft SP will also need to be revised extensively. Only then can recommendations related to the level of the primary or secondary standard be considered.

## **Comments on Chapter 2 – Air Quality Characterization**

Chapter 2 begins with the statement that the chapter summarizes the origin and status of ozone concentrations in the ambient air. Section 2.2 begins with the statement that ozone is formed in polluted areas from atmospheric reactions of two main precursor classes, volatile organic compounds and nitrogen oxides. Only later it is noted that ozone is found everywhere in the atmosphere and still later the various natural processes that contribute to ozone are noted. While EPA's focus is on ozone from man-made precursors, the Chapter needs to do a better job of putting the ozone issue into context with the overall atmospheric levels of ozone and the mix of natural and man-made components. A national ozone standard applies everywhere in the U. S., not just in and around major urban centers during the ozone season.

It is important at the outset for the SP to note that ozone is both a natural constituent of the atmosphere (through transport from the stratosphere) and a secondary pollutant formed from atmospheric reactions of organic and nitrogen oxide precursors. Major sources of precursors include both natural and man-made processes. In addition to combustion, lightning and microbial action in soils are major NO<sub>x</sub> sources globally and in the U. S. In addition, agricultural activities are a major source of NO<sub>x</sub> and ammonia (some portion of which reacts in the atmosphere to form NO<sub>x</sub>) as well as organic precursors. Organic precursors include both man-made and biogenic emissions.

While Chapter 2 appropriately includes a short summary of the chemical and physical processes that determine ozone concentrations in the atmosphere, the chapter should focus on the concentrations, patterns and exposures that are relevant to understanding and interpreting the health and welfare effects data and ultimately inform the choices of appropriate standards to protect the public health and welfare. In its current form, it is

---

<sup>1</sup> J. M. Heuss, Comments on the First External Review Draft of "Air Quality Criteria for Ozone and Related Photochemical Oxidants," Air Improvement Resource, Inc. report prepared for General Motors Corporation, April 26, 2005; J. M. Heuss, Comments on the Second External Review Draft of "Air Quality Criteria for Ozone and Related Photochemical Oxidants," Air Improvement Resource, Inc. report prepared for General Motors Corporation, November 30, 2005.

deficient. As discussed below, the draft fails to provide the reader with sufficient information concerning the levels of ozone relative to the current standard or alternatives that will be considered and fails to adequately characterize policy relevant background for later use in the risk assessment.

#### **Section 2.2.4 Precursors, sources and emissions**

This section focuses on man-made emissions of VOC and NO<sub>x</sub> allocating four pages to a detailed summary of VOC and NO<sub>x</sub> emission estimates for various source categories from 1970 to 2004 reported to 5 significant figures. The relevant information from these tables could be summarized in a few sentences as follows:

EPA's latest emissions and air quality trends report<sup>2</sup> documents that emissions of the three major precursor categories have all been reduced substantially. EPA reports that between 1970 and 2002, nationwide emissions of VOC have been reduced 51%, CO has been reduced 48 % and NO<sub>x</sub> has been reduced 17 %. For the 20- year period 1983 to 2002, EPA reports a 40 % reduction in VOC a 41 % reduction in CO and a 15 % reduction in NO<sub>x</sub> emissions. The ambient air quality improvements that have been measured indicate that there is a greater reduction in measured air quality than in nationwide emissions. This is due to spatial growth in population and emissions as cities have grown and spread out. For example, CO emissions were reduced by 41 % from 1983-2002, but ambient CO levels were reduced by 65 %. Similarly for NO<sub>x</sub>, a 15 % reduction in NO<sub>x</sub> emissions was accompanied by a 21 % reduction in NO<sub>2</sub> levels. Although there is not a routine VOC monitoring network, the reduction in ambient man-made VOC precursors should be comparable to that of CO. Monitoring of individual air toxic compounds that are also ozone precursors confirms that there has been a major reduction in ambient levels of man-made VOC.

In contrast to the detailed coverage of man-made emissions, a critical issue that is not well addressed in the SP is what portion of the ozone measured in different locations is controllable through control of U. S. man-made emissions. Annex 2 of the second draft CD contains many specific facts that are relevant to understanding the processes and sources that contribute to background ozone. For example, lightning is a source of NO<sub>x</sub> that contributes to both the planetary boundary layer and the free troposphere. The annex on page AX2-125 notes that the magnitude of the lightning source is roughly 10 % of global NO<sub>x</sub> emissions but that estimates of the magnitude vary by a factor of ten. Microbial action in soil is another major source of natural NO<sub>x</sub> that peaks in the summertime. The EPA estimates of NO<sub>x</sub> from lightning and soils are given in Table AX2-5 indicate that natural NO<sub>x</sub> in the U. S. is about 13 % as large as the current man-made sources. In addition, the fact that biogenic VOC is greater than anthropogenic VOC on a regional and global basis is discussed on page 2-20 of the CD.

This section should include estimates of the relative precursor emissions from man-made and natural processes in the U. S. Since agricultural activities do not appear to be

---

<sup>2</sup> U. S Environmental Protection Agency, National Air Quality and Emissions Trends Report, 2003

included in Tables 2-1 and 2-2, staff should also provide estimates of the VOC and NO<sub>x</sub> emissions from agricultural activities since they play a role in uncontrollable background.

### **Section 2.3 Data sources**

The available information on the precision and accuracy of ozone monitoring is relevant to the question of how precise an air quality standard can be and whether the current round-off procedure is appropriate. At the December 2005 CASAC meeting the question of precision of ozone monitoring was raised and California's new 8-hour ozone standard was referred to. Although California has specified an 8-hour ozone standard to the third decimal place in ppm, the California staff acknowledges that the third decimal place is an uncertain digit.<sup>3</sup> In fact, staff indicates that the average bias from audits of 137 monitors in 2003 was - 2 % with a standard deviation of 4.3 %. The staff interprets this to mean that ozone concentrations are uncertain in the third decimal place, to the degree of plus or minus 0.003 ppm. Since one standard deviation was 4.3 %, two standard deviations would be 8.6 % and the interpretation would be that 95 % of the time the reported ozone at 0.070 ppm is within plus or minus 0.006 ppm of the true concentration.

The SP discusses precision in the national network but the presentation is not particularly clear. The text indicates that the overall mean precision was 3 % for 2002 and, when special purpose monitors were also considered, the average upper bound precision was 2.9 %. (page 2-10) Additional detail should be provided so that the precision of the current monitoring network is clearly defined for the reader.

The California and EPA estimates of precision of plus or minus several percent are consistent with the precision reported for the CASTNET sites that 98 % of the audits were within plus or minus 10 %. (page 2-10) Given this level of precision, the current round-off procedure is appropriate.

There is also a question of interferences in the UV monitors that has been raised by the American Petroleum Institute.<sup>4</sup> We urge staff to fully explore the extent and ramifications of such interferences and modify the ozone network to assure that interferences do not influence the attainment/nonattainment designations for ozone.

### **Section 2.5 Characterization of ground-level ozone concentrations**

This section includes one map of the locations of ozone monitors (Figure 2-1) and 18 other maps and figures displaying ozone data in various ways. Figures 2-2 and 2-3 give ozone data on a county basis for the metrics of the 8-hour and 1-hour standards. There are two issues with this choice of data presentation. First, there is sufficient data variation within counties that it is misleading as to spatial extent and the attainment/nonattainment status of each monitor. Second, the data is presented in only three bins which masks the range in the data.

---

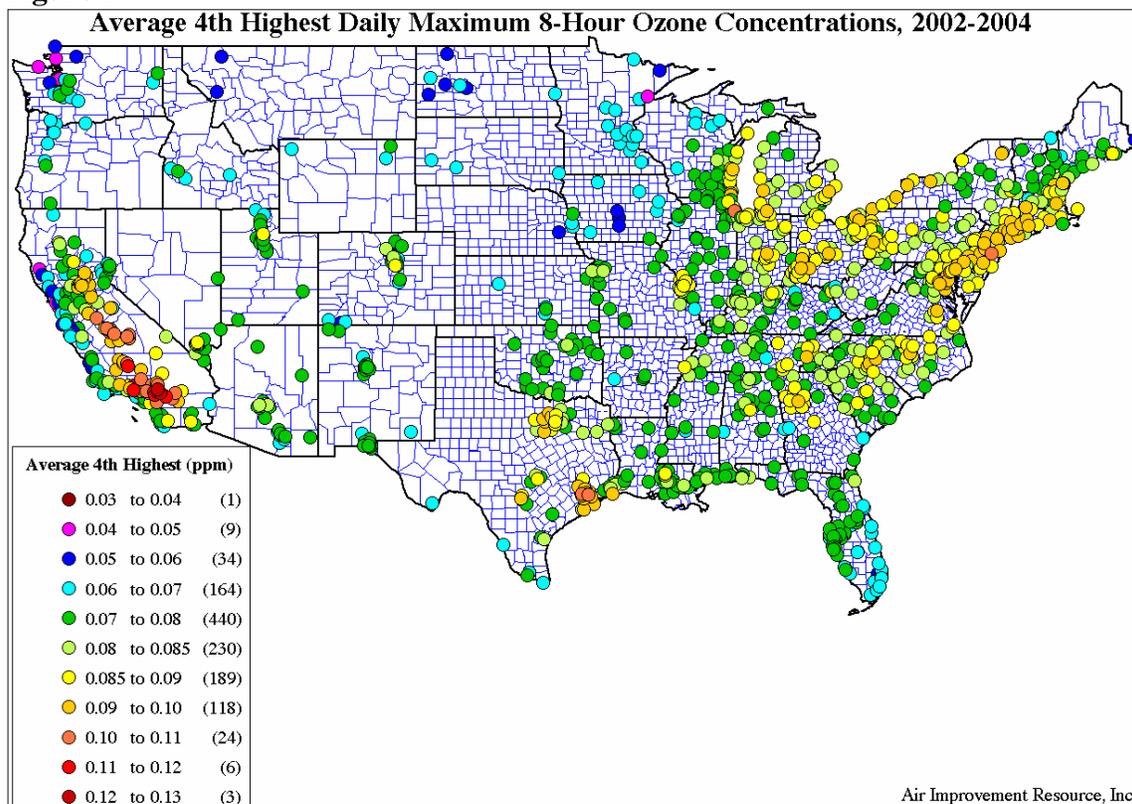
<sup>3</sup> California Air Resources Board Staff Report, Review of the California Ambient Air Quality Standard for Ozone, Volume II, March 11, 2005, page 6-3.

<sup>4</sup> See Attachment C of American Petroleum Institute comments on the ozone CD, November 30, 2005.

In our April 2005 comments on the first draft CD, we provided maps that evaluated the yearly 8-hour maxima in the western U. S. The figures showed that almost all the rural and remote sites in the western half of the nation have yearly 8-hour maxima that fall in the range of 0.06 to 0.08 ppm for 2001 through 2003.

We recommend that a variation of this data presentation be included in the SP. The mean of all hours, and the design value concentrations for the 8-hour and 1-hour standards should be plotted for each site in the database. By doing this, the reader will be able to evaluate the spatial variation in several metrics that are relevant to interpreting the health data and to understanding the background. The mean of all hours metric is important because it bears on the interpretation of the epidemiologic studies. The 4<sup>th</sup> highest 8-hour average metric is important because it provides perspective related to the current form of the 8-hour standard. An example of such a map for the current form of the 8-hour standard is shown below in Figure 1. The map displays the design values (calculated using EPA's methodology) based on 2002 to 2004 data for all the sites in the Agency's database. The data is presented in 10 ppb bins except for the 0.08 to 0.09 ppm bin which is split in half to enable the reader to distinguish which sites do not attain the 0.08 ppm standard.

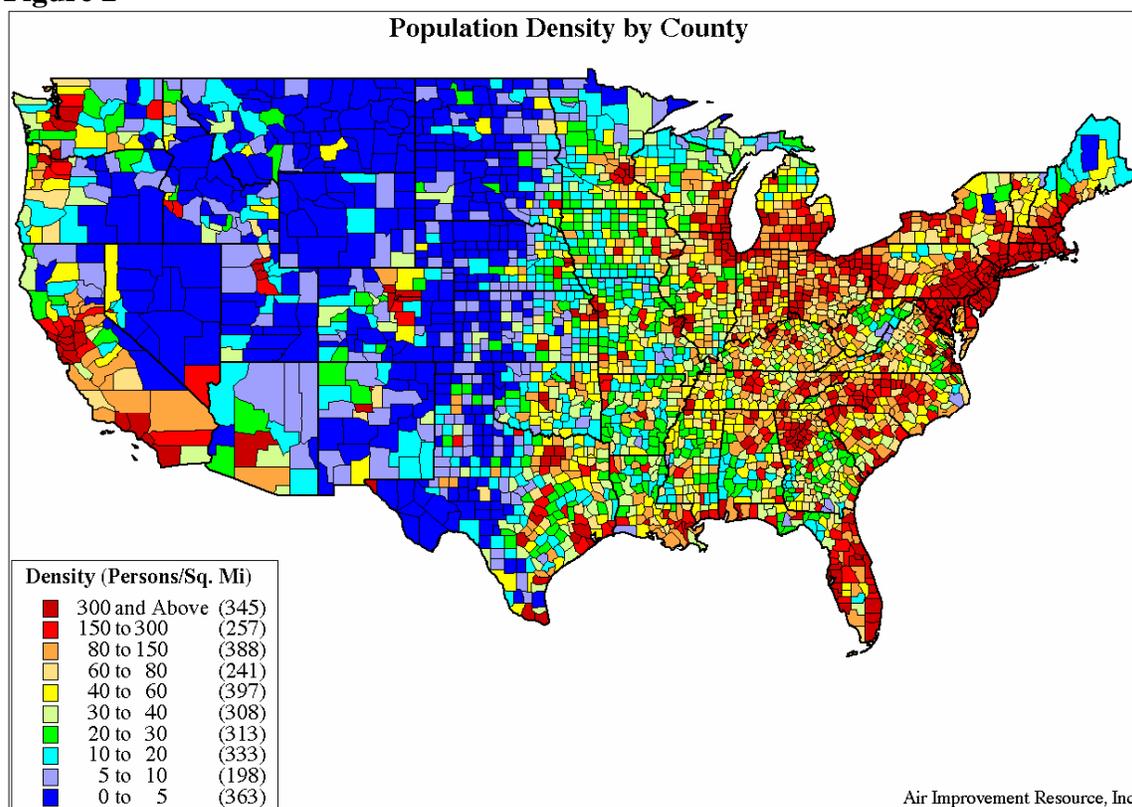
**Figure 1**



A point made on page 3-9 of the 2<sup>nd</sup> draft CD is that it is hard to find sites in the eastern U. S. that are not influenced by transport from urban areas. That is correct, but that is why a careful evaluation of the ozone levels in the western U. S. is, by way of contrast, extremely important. The population and emissions density in most of the western U. S. is only a few percent of that in the eastern U. S. Therefore, an evaluation of the peak 8-hour values at western monitoring sites can offer one way to estimate PRB and may provide a practical limit to how low an ozone standard can be before it conflicts with background.

Figure 2 is a map displaying the population density by county for the U. S. It shows that large areas of the western U. S. have population densities between 0 and 10 persons per square mile. This contrasts to the population and emission densities in the major urban areas of the U. S. For example, the population densities in the top 50 counties in the nation are all above 2,000 persons/square mile and for the top 10 counties are all above 8,000 persons per square mile. The densest-populated state is New Jersey with an average population density of about 1,000 per square mile. Since the distribution of population provides a reasonable surrogate for the distribution of man-made emissions, the large difference in population and emissions between the east and west can be used to evaluate what ozone levels would exist with almost total elimination of man-made emissions. It is as though we run a large experiment each day with many times more emissions in the eastern U. S. than in the interior western states.

**Figure 2**

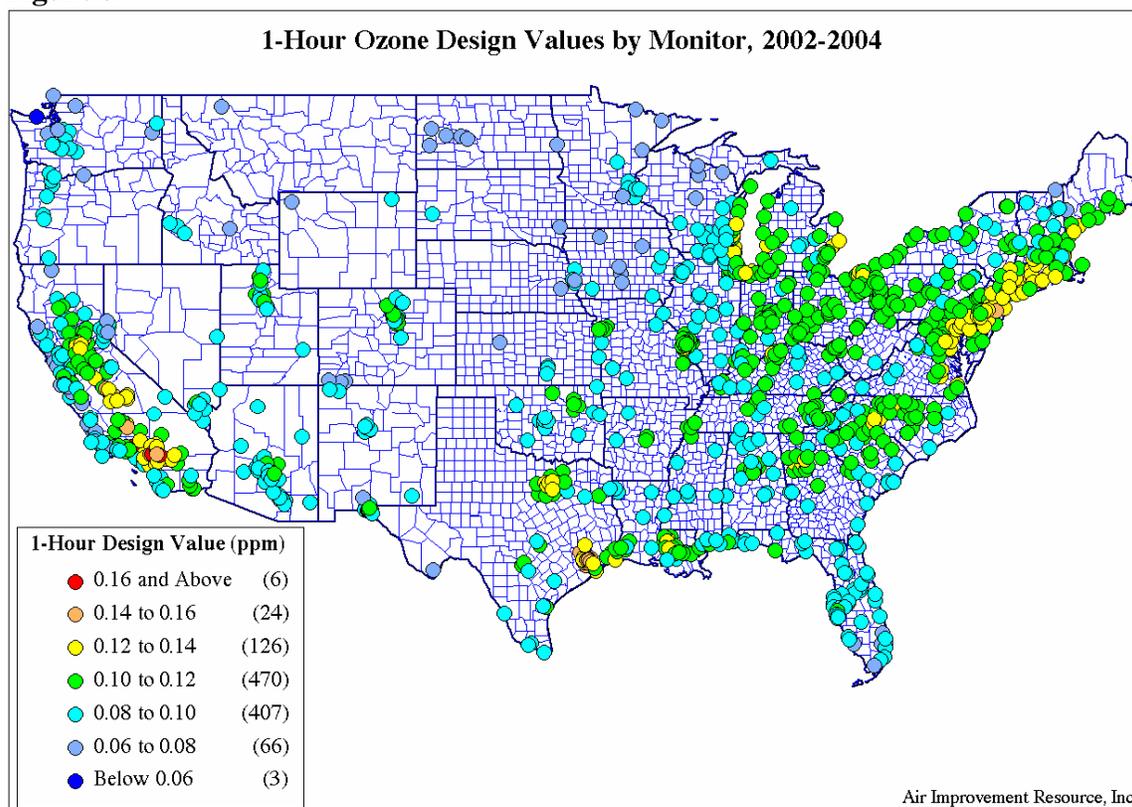


As shown in Figure 1, there are many western sites with 4<sup>th</sup> highest 8-hour averages that are in the range of 0.06 to 0.07 ppm and many in the range of 0.07 to 0.08 ppm. The sites with the lowest ozone design values tend to be in central cities along the west coast (where ozone is suppressed by the presence of NO<sub>x</sub>), at low elevation coastal sites along the west coast (where vertical mixing is suppressed due to prevailing subsidence), and in remote sites in the northern tier of western states. The concentrations at low elevation coastal sites are low since there are no biogenic precursor emissions over the ocean and the prevailing subsidence limits the contribution from ozone transported aloft from Asia. At elevated sites along the Western tier of states, such as at Mt. Lassen volcanic park in northern California, there are substantially higher extreme values and evidence that a substantial number of these are from Asian transport.

Based on current ozone levels, an 8-hour standard set as low as 0.06 ppm would put virtually the entire country out of attainment leaving essentially no room for ozone from human habitation.

Figure 3 displays the 2002-2004 peak 1-hour ozone levels across the country, and Figure 4 displays the ozone season mean concentrations. Inspection of Figures 1, 3, and 4, in relation to the population density in Figure 2, shows that the highest peak 8-hour levels and the peak 1-hour levels are located in and immediately downwind of the major populated areas of the country, as expected. In the remote areas of the west, removed from significant emission sources, the peak 8-hour levels are in the range of 0.05 to 0.08 and the peak 1-hour levels are in the range of 0.06 to 0.10 ppm. When compared to the population density map, the peak ozone values reflect the presence of a substantial contribution from man-made ozone in and immediately downwind of populated areas on top of a significant background of ozone.

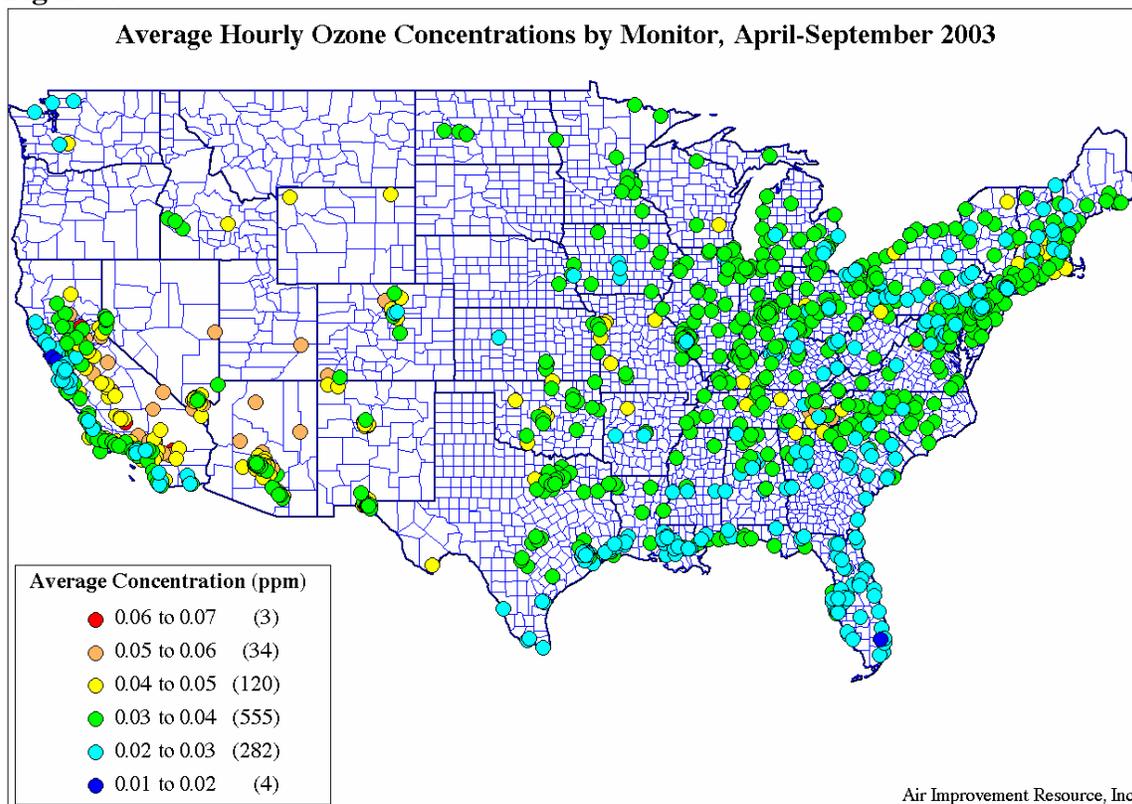
**Figure 3**



In contrast, the mean ozone concentration shown in Figure 4 displays a different behavior. The vast bulk of sites are in the range of 0.02 to 0.04 ppm. Higher levels are found inland in California and at western sites. The similarity of mean ozone across the country has several implications. First, the day-in-day-out ozone exposures of the population during the ozone season are low and similar in populated and remote areas. Since human exposures average half or less of these values because people spend about 90 % of their time indoors, the low long-term exposures provide an explanation for lack of a chronic effects signal as documented in the CD. Second, the day-in day-out exposures of the population that are implicated as potentially causing serious health outcomes in the epidemiology studies are generally very low. This provides a serious

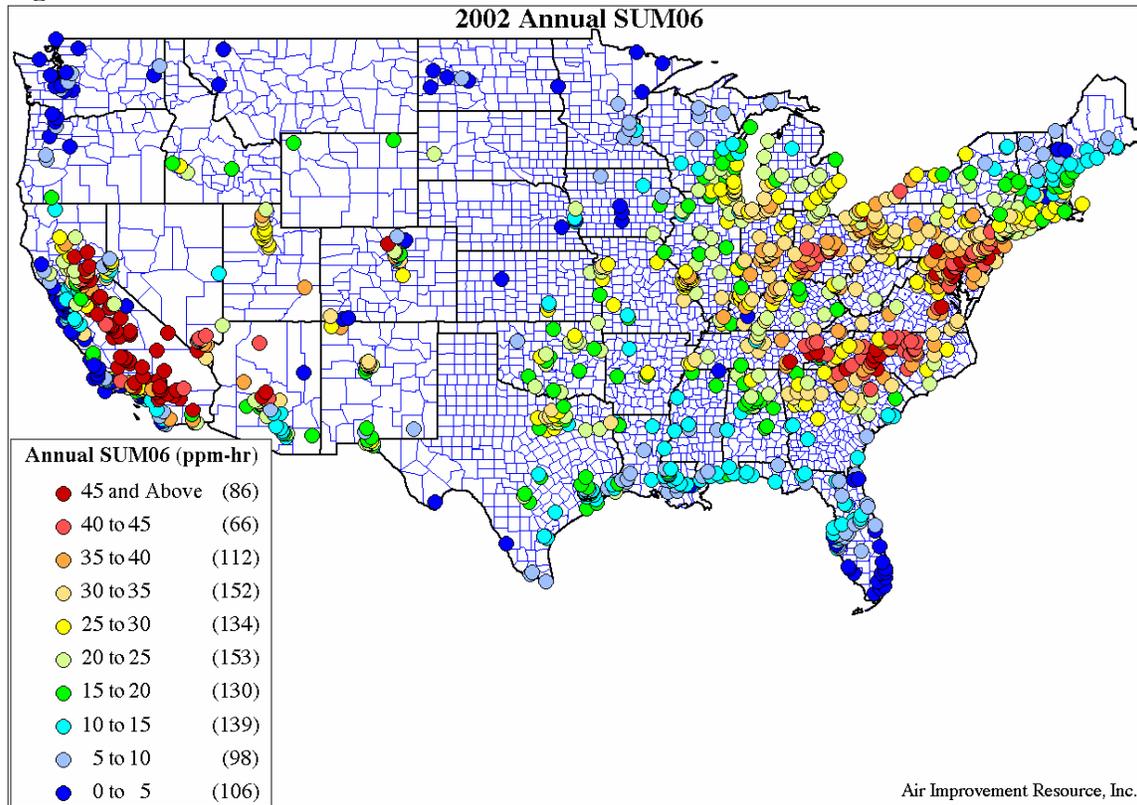
challenge for establishing biologic plausibility of the reported epidemiologic associations with serious health outcomes. Third, the similarity of mean ozone exposures in areas with population and emissions densities that vary by a factor of 100 or more indicates that even virtual elimination of man-made emissions will not change these exposures to a significant degree.

**Figure 4**



Figures 2-4 through 2-7 of the SP display ozone metrics that may be considered for the secondary standard. One of these is the highest three-month SUM06 which was proposed but not adopted as the metric for the secondary standard in the 1996/7 review. As for the maps presenting peak 8-hour and 1-hour values in the SP, these maps do not provide sufficient detail. Figures 5 through 7 below plot the three-month SUM06 values (in ppm-hr) for the entire EPA database using more bins for the data for the years 2002, 2003, and 2004. These figures show particularly high SUM06 values in and directly downwind of Los Angeles and the Central Valley of California as well as wide variations from year to year in the Eastern U. S. There is also a North to South gradient in the inland Western states. As for the peak 8-hour and 1-hour levels, the overall pattern suggests both a contribution from U. S. man-made emissions and a background component.

**Figure 5**



**Figure 6**

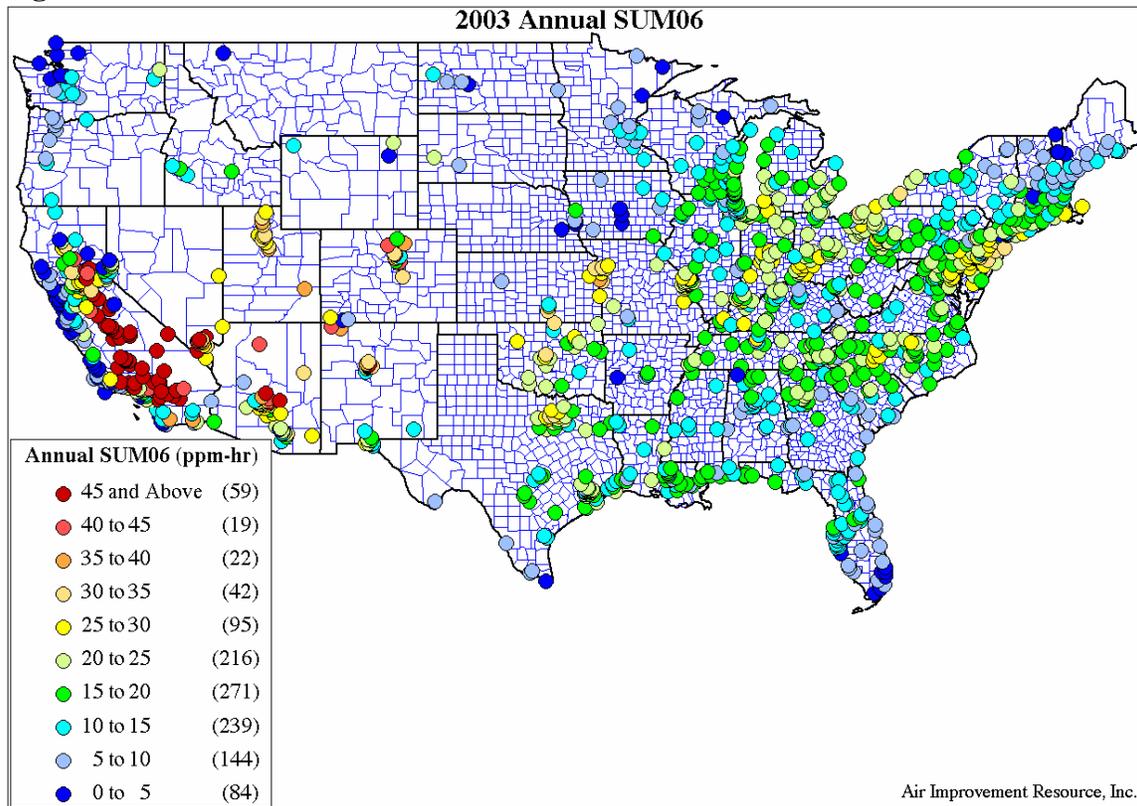
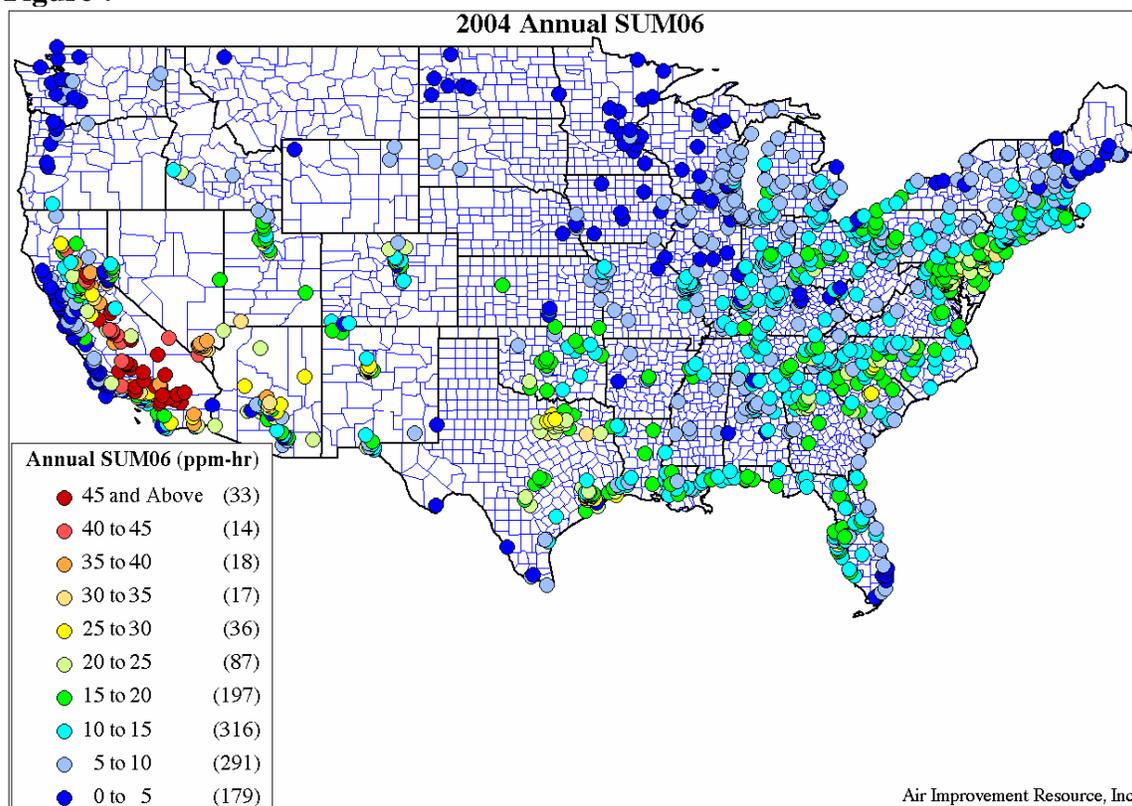
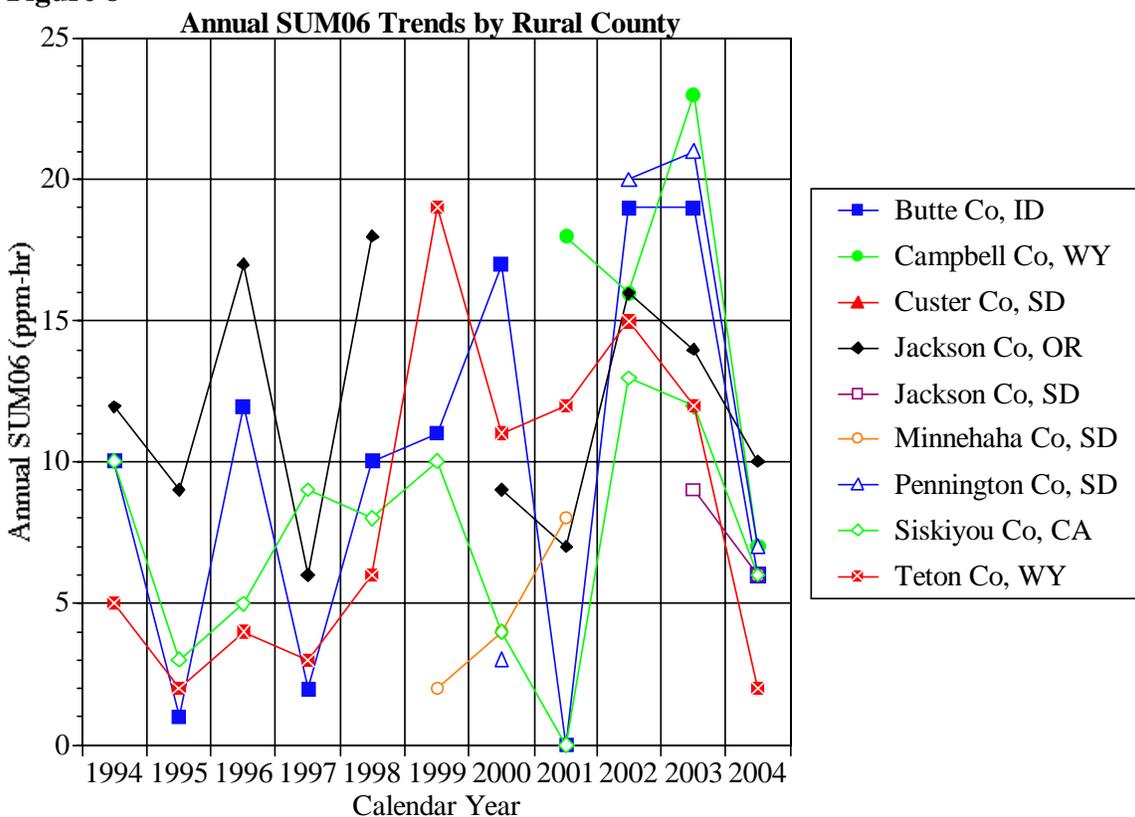


Figure 7



To further evaluate the background component, we plotted the longer term SUM06 data for a series of sites across the western U. S., from the California/Oregon border across Idaho, Wyoming and into South Dakota, in Figure 8. All of these sites are substantially removed from major emission sources and have SUM06 values generally in the range of 0 to 20. Note also that there are wide variations from year to year. There are several important implications from this data. First, even these sites removed from obvious source areas have a substantial number of hours each year at or above 0.06 ppm, the order of 100 to 200 hours per year. In fact, all these sites have 8-hour design values of 0.06 or greater in 2002-2004 as shown in Figure 1. Second, the SUM06 metric varies substantially from year to year even in rural and remote areas. This means that if a SUM06 standard were set using only one year's data, areas would bounce in and out of attainment causing havoc with implementation planning efforts. Third, the presence of a substantial amount of SUM06 across the entire country indicates that SUM06 is partially confounded by background.

Figure 8



### Section 2.7 Background levels

Section 2.7 begins with a definition of policy relevant background as those concentrations that would result in the United States in the absence of anthropogenic emissions in North America. As we have argued in comments on the CD, the definition of policy relevant background (PRB) in the CD and SP is flawed. It omits consideration of the emissions from agricultural activities and it omits consideration of the contribution to ozone from anthropogenic emissions in Mexico and Canada. While there can be reductions in emissions from agricultural activities, it would be impossible to drive them to zero. Surely staff does not contemplate the cessation of the practice of growing crops and raising animals in the U. S. The impact of Canadian and Mexican emissions is also problematic because, even though the U. S. may have treaties with these countries, zero anthropogenic emissions in these countries is not a possible future outcome. At a minimum, the Agency should evaluate the impact of these two major sources – agricultural activities and neighboring country’s emissions - on PRB

The SP notes that because of long-range transport from anthropogenic source regions within North America, estimates of PRB cannot be derived solely from measurements of ozone. However, it goes on to state that PRB estimates must be based on modeling. This is not true. While transport of plumes from urban areas is well documented in the

literature, the CD offers no observation-based analyses to demonstrate that the many instances of ozone greater than 0.06 ppm that are measured at remote monitoring sites are all caused by transport from urban areas. As documented in Annexes 2 and 3 of the CD, there are still major uncertainties and limitations involved in using atmospheric models. Therefore, both observations and models should be used to estimate PRB. In addition, PRB needs to be evaluated for various ozone metrics and exposure situations. If extreme value forms of the ozone standard are to be evaluated in the Staff Paper, then estimates of the extremes of PRB should be included in the SP.

The SP relies solely on the Fiore et al. 2003 global modeling study. However, the CD notes several important limitations of that study and the need for additional work to evaluate PRB with an ensemble of models comparing the model results with observations, including models with greater resolution to understand variability on shorter time scales and variability due to processes that are not captured in the global models. (2nd draft CD at Page 3-52) Until this is done, a range of background estimates from models and observations should be used. For the health and welfare risk assessments, several alternative background assumptions should be used. They could include a constant 40 ppb background, scaling the seasonal pattern from the Trinidad Head data, and a probabilistic formulation that provides yearly peak 8-hour values in the range of 60 to 75 ppb.

The CD notes that one of the major issues with global models is how they parameterize or characterize stratospheric-tropospheric exchange and lightning flash rates (see pages ax2-126 and ax2-110 of the CD). Many of the scientific issues with the chemical transport models are discussed in section AX2.5. These include the difficulty of any model to capture troposphere folding events, problems with the treatment of clouds that can influence the ozone formation rate by 50 %, and the importance of horizontal grid size relative to mixing. In addition, the chemical mechanisms used are highly simplified compared to the many reactions that are known to occur in the atmosphere.

The CD lists 14 global models in Table AX2-4 and presents comparisons of their predictions with observations in Figures AX2-19 and 20. It is clear from the figures that there is substantial variation in global model results. While their performance is reasonable, no one model is a perfect reflection of reality. The text in the CD notes that the largest discrepancies in the models reflect differences in their handling of stratospheric-tropospheric exchange. While present global models can reproduce much of the variability in lower troposphere ozone, they are not able to simulate many of the processes that are described in sections AX2.3.1 and 2.3.2. Many of the processes occur at scales that cannot be resolved by current models thus limiting the model's ability to attribute trace gases to their sources. (page AX2-69).

The CD also includes discussion of several large experimental programs that document long range transport of ozone and precursors, the presence of a wide variety of air masses with different chemical content, and many examples of mixing of anthropogenic-influenced and stratospheric-influenced air as well as the existence of layers of air with

very different chemical composition in close proximity to one another. (see Section AX2.3.4)

All these scientific findings in the CD bear on the issue of policy relevant background that is discussed in the SP. They provide evidence for mechanisms that can result in elevated ozone from stratospheric-influenced air and document the limitations of the Fiore et al. 2003 modeling study at simulating these mechanisms. In order to evaluate the range and variability in PRB, the SP needs to consider both observations and models. Since the ozone standards have historically been defined as yearly extremes of daily maxima, the SP should evaluate PRB for these statistical forms as well as for mean conditions. When the agency estimated background ozone during review of the federal ozone standard in 1996/7, the agency's Staff Paper concluded:

“...a reasonable estimate of the background O<sub>3</sub> concentrations near sea level in the U. S. for a 1-hour daily maximum during the summer is usually in the range of 0.03 to 0.05 ppm. At clean sites in the western U. S., the maximum annual hourly values are in the range of 0.06 to 0.075 ppm.” (1996 OAQPS Staff Paper at page 20)

There is extensive discussion of mean PRB ozone in the CD. The 1996 CD presented a range of from 0.02 to 0.035 ppm. In the current draft CD, estimates that range from 0.015 to 0.045 are included in Table 3-2 for various regional or seasonal means. Annual or seasonal mean levels of background are an important consideration when it comes to evaluating human exposure from background ozone and in interpreting the health studies. However, for control considerations, because of the form of the ozone standard, the background related to extremes of the distribution of daily maxima as well as the means need to be evaluated. In addition, the issue is not only what the background is on days of maximum photochemical potential, but also what the background is on days with the highest background.

It is particularly important that the qualifications and limitations concerning background in the CD be carried over to the Staff Paper. As we pointed out in our earlier comments, the Fiore et al. 2003 modeling study used a global transport model that was not designed to specifically address the components of background in the U. S. in relation to extreme value standards. Any global model contains many assumptions and simplifications that simply cannot be fully evaluated. The GEOS-CHEM model is but one of a number of such models as demonstrated in Chapter 2 of the CD.

We also pointed out that Fusco and Logan 2003<sup>5</sup> evaluated the GEOS-CHEM model and report that the model estimates somewhat higher production and loss rates of ozone than other chemical transport models, as much as 15 to 30 %. Since the net photochemical production of ozone is determined by the difference between these two large numbers (a large chemical source term and a large chemical sink term), the net production cannot be precisely determined. They note that differences in modeled photochemical production

---

<sup>5</sup> A. Fusco and J. Logan, “Analysis of 1970-1995 trends in tropospheric ozone at Northern Hemisphere midlatitudes with the GEOS-CHEM model,” *J. Geophys. Res.*, **108**, No. D15, 4449 (2003).

and loss rates affect the relative importance of the stratospheric source giving examples of other models that indicate a much larger role for the stratospheric source in summer and in winter. Adding to the complexity is that assumptions have to be made about the cross-tropopause flux of ozone and ozone deposition at the surface, quantities that each have significant uncertainty, too. There is substantial disagreement over key factors such as the magnitude of U. S. biogenic VOC emissions (uncertain by a factor of 2 or 3) and natural NO<sub>x</sub> emissions from soil and lightning. There is also uncertainty due to the chemical formulation which is only a simplified approximation of the chemical processes actually occurring in the atmosphere.

The large grid size in the model is also problematic for getting the regional and local ozone production right. The large grids mix or smear precursor emissions from areas hundreds of kilometers on a side together in ways that do not occur in the real world. Since ozone formation is complex and non-linear, the model's mixing of emissions may result in bias. This is one of the reasons why the first draft CD indicated that "local maxima and minima are difficult to reproduce with global models since processes are averaged over an entire model grid cell." (Page AX2-116) Liang and Jacobsen<sup>6</sup> have shown that integrated ozone production may be over- or under-predicted when coarsely resolved models blend air masses of different origin. Over-predictions by as much as 60 % were found for coarse-model grid cells exposed to different air masses.

During the December 2005 CASAC meeting Dr. Cowling asked a perceptive question, "How does the GEOS-CHEM model perform for the ozone precursors?" Because the model mixes emissions in large grids, it clearly does not simulate the concentrations of precursors correctly on the scale that influences daily ozone peaks. So even if the model predicts the correct ozone, it is doing it for the wrong reason.

The first draft CD also indicated that the simulation of stratospheric intrusions is notoriously difficult in global models (AX3-157). This arises because stratospheric-tropospheric exchange is an intermittent process, a series of strong exchange events separated by more quiescent periods.<sup>7</sup> The events insert stratospheric air high in ozone into layers or filaments in the troposphere. Only rarely do these intrusions penetrate to ground level intact, but they intermittently insert high concentrations of ozone into the troposphere much more often. The layers or filaments of stratospheric ozone eventually mix with surrounding air in ways which are not yet fully understood, influencing ozone levels, photochemical activity and lifetimes of various species. The CD acknowledges that coarse-grid models cannot capture the structure and implications of these phenomena (see CD at pages AX3-144, -146, and -148). Since the current ozone standards (1-hour and 8-hour) are short-term concentrations with extreme value statistical forms, the question comes down to how well global models can simulate the impact of background and anthropogenic emissions on local short-term extreme events. The current draft CD indicates many sources of uncertainty and variability in addressing that task. For all

---

<sup>6</sup> J. Liang and M. Jacobson, "Effects of subgrid segregation on ozone production efficiency in a chemical model," *Atmospheric Environment*, **34**, pages 2975-2982 (2000).

<sup>7</sup> A. Stohl et al., "Stratospheric-tropospheric exchange: A review, and what we have learned from STACCATO," *J. Geophys. Res.*, **108** (D12), page 8516 (2003).

these reasons, the Agency should not rely solely on this one modeling study to define policy relevant background.

### **Comments on Chapter 3 –Policy-Relevant Assessment of Health Effects Evidence**

The Chapter significantly overstates the consistency, plausibility and coherence of the epidemiological evidence. Although there is now a distinctly larger data base of epidemiological studies related to ozone, a number of critical findings in the epidemiological literature - some of which are noted in the CD and some of which are not - are omitted or downplayed in the Chapter. The Chapter begins by noting that the last ozone review focused substantially on controlled exposure studies and that the current review emphasizes a large number of epidemiological studies. The text describes the highlights of this new evidence as (1) large multi-city studies that examine the acute effects of ozone on mortality that provide much more robust and credible information than was available for the last review, (2) new acute studies that offer added evidence of lung function decrements and respiratory symptoms in exercising healthy and asthmatic subjects as well as evidence on new health endpoints such as school absences and cardiac physiologic endpoints, and (3) several new studies examining associations between ozone and emergency department visits for respiratory diseases and a significant effect on respiratory hospital admissions.

As documented below, the SP overestimates the magnitude and consistency and underestimates the uncertainty in the results of acute epidemiologic studies, especially for symptoms, school absences, hospital admissions, and mortality. While there is a larger database than in 1996, there is an implausibly wide range of results in individual cities ranging from positive to negative in systematic analyses, and issues of model selection and publication bias cloud the interpretation of the data.

Ozone is a well known respiratory irritant and, for or some health endpoints, such as temporary decrements in the performance of lung function tests and symptoms, there is sufficient information from controlled studies to anticipate that there may be corroboration in observational studies. For other endpoints, such as mortality and hospital admissions, there are many other known causes and the issue for this Chapter is to address the question whether ozone in particular is causally related to these endpoints.

#### **Section 3.2 Mechanisms**

There is now a large body of controlled studies that provide information on ozone effects and the detailed mechanisms that cause those effects. The determination of the medical significance of single or repeated exposures to ozone at various levels and the plausibility of various health effects implicated by epidemiologic studies needs to be evaluated with consideration of the body's defense mechanisms against an irritating gas such as ozone. These defenses include the presence of antioxidants in the epithelial lining fluid, a neurally-mediated reflex action to inhibit maximum inspiration, and initiation of immune

system responses that resolve within a few days and at higher doses tissue damage and repair mechanisms that resolve over longer time periods.

The controlled studies in the CD clearly demonstrate that the mechanisms through which ozone exerts toxic effects are non-linear or threshold-based. The fact that lesser effects are observed earlier and at lower exposures than more severe health effects is an important consideration when it comes to evaluating the plausibility of the epidemiologic associations reported in the literature. This section needs to note the non-linear or threshold-based nature of the mechanisms discussed.

The text also notes that, in the past, little information was available to help explain biological mechanisms which linked ozone to premature mortality or cardiovascular effects. This is true. It is also true that, despite the speculation in section 3.2.2, there is still no credible explanation for how the low ozone exposures experienced on a day-to-day basis by the population could cause premature mortality or cardiovascular effects. As several CASAC panelists noted during the December 6-8, 2005 meeting, the material in the second draft CD and the first draft SP over-interpreted and overstated the case for cardiovascular effects. The limited database reporting small changes in cardiac parameters associated with ozone or other pollutants in observational studies could easily be confounded by a host of factors, and are not found consistently for any specific pollutant. In addition, the changes are small and within normal background noise due to changes in day-to-day activities and are too small to be of clinical significance.

The data from controlled studies such as Gong et al. 1998 do not suggest any effects from typical day-to-day ozone exposures. Gong et al.<sup>8</sup> exposed subjects to 0.3 ppm ozone for 3 hours with exercise and found little direct effect of ozone on the heart or vasculature for hypertensive or healthy subjects even though many pertinent variables (such as cardiac index, ventricular performance, pulmonary artery pressure, pulmonary and systemic vascular resistances, ECG, serum cardiac enzymes, plasma catecholamines, and atrial natriuretic factor, for example) were evaluated. Gong concluded that they did not find convincing evidence of significant direct cardiovascular effects from ozone exposure, analogous to effects reported in laboratory animals. The small indirect effect they report, impaired gas exchange, can be explained as an effect secondary to the respiratory effects of the high ozone exposures. The extremely low personal exposures of the population on a day-to-day basis are not expected to cause any cardiovascular effects. The text in Section 3.2.2 should be changed to acknowledge that we still do not know how ozone could cause serious cardiovascular effects, or premature mortality in general.

### **Section 3.3 Nature of effects**

#### **Section 3.3.1.1 Emergency department visits/hospital admissions for respiratory causes**

---

<sup>8</sup> A full report of the Gong et al. 1998 study is available as a report to the California Air Resources Board, Final Report Contract No. 93-327, May 2000.

The SP acknowledges, as the CD notes, that the body of studies is larger now than it was in 1996/97, but that, for emergency room visits, the data is now inconclusive as regards an ozone association. There is a very wide range of associations in the literature. For asthma-related emergency room visits, the range is from -20 % to + 80 % per standard ozone increment as shown in Figure 7-8 of the CD. For total respiratory hospital admissions, the range is from - 15 % to + 15 % per standard ozone increment as shown in Figure 7-9.

While there are statistically significant associations reported for respiratory hospital admissions in a variety of studies, there are also studies where there was no association or a negative association. Most studies evaluated associations with either total respiratory diseases or with asthma. Total respiratory causes include asthma, pneumonia, bronchitis, emphysema, respiratory cancers, and upper and lower respiratory infections such as influenza. Chronic bronchitis and emphysema are often combined to define chronic obstructive pulmonary disease (COPD). In the recent California review, the staff acknowledged that “no clear pattern is evident regarding associations of ozone with specific respiratory disease outcomes.”<sup>9</sup>

In addition, the CD notes several key studies from California that reported no association even though it is a high ozone area. For example, Nauenberg and Basu 1999 evaluated associations between asthma-related hospital admissions and exposure to PM10 and ozone. They found no association between asthma and ozone even when seasonal analyses were performed.

In a study that is not included in the CD, Van Den Eden et al.<sup>10</sup> carried out a study of hospital admissions in the Central Valley of California using the GAM approach. Ozone had a negative association, that is it was associated with a decreased risk of both acute respiratory and chronic respiratory admissions. Although carried out under contract to the California Air Resources Board, this study has not been published in the open literature. Although not published, this study reinforces the conclusion that there is a lack of ozone effects observed in high ozone areas of California. It also provides an example of the subtle impact of publication bias by which no effect studies or studies that found actually negative associations are generally not published.

There are three general categories of time-series studies in the literature cited in the CD. Because of the interest in PM health effects for the past decade, some studies focused on PM associations with health endpoints but included ozone in the analysis. A second category focused on ozone associations and may or may not have included other pollutants. A third category of studies evaluated a suite of pollutants including ozone.

No effect or negative effect studies do appear in the literature from the first and third categories, but generally do not from the second category. In all three categories, there is

---

<sup>9</sup> California Air Resources Board Staff Report, Review of the California Ambient Air Quality Standard for Ozone, June 2004.

<sup>10</sup> Van Den Eden et al., 2002, Central Valley hospital admissions study for ARB, Final Report ARB Contract 97-303.

almost always some positive association with at least one pollutant reported, even if multiple comparisons were carried out and the authors indicate the positive association they report may well have occurred due to chance. The overall result is a wide range of associations for any of the health endpoints evaluated in the SP for ozone as well as for all the other criteria pollutants that are routinely monitored.

The presence and importance of publication bias in air pollution epidemiology was noted in both oral public comments and by CASAC panelists during the December 2005 CASAC meeting. However, the SP essentially ignores the issue. From the wide range of associations in the literature, it is possible to select some to implicate each pollutant in health effects. Depending on the selection and the way the data is presented, the appearance of generally consistent associations in single-pollutant models can be shown.

Although confounding by co-pollutants is discussed in the CD, the discussion is incomplete. Examples of confounding by PM are discussed but confounding by other pollutants is not. Since associations of respiratory admissions have been reported with other gases as well as with PM and ozone in single-pollutant models, a more complete evaluation of the extent of confounding by other pollutants is necessary. While there are some examples of situations where ozone associations were robust when other pollutants were considered, there are also other examples where it was not. Therefore, the staff conclusion that positive and robust associations were found between ambient ozone and hospital admissions (page 3-8) overstates the consistency of the data and should be highly qualified.

For hospital admissions, there are many issues with the studies the text focuses on. The largest study to date is Burnett et al. 1997. Burnett et al. evaluated the association of hospital admissions for respiratory causes from 16 cities across Canada over a ten-year period. However, it is not clear whether the authors evaluated only unscheduled visits as recommended in the introduction to this section of the SP. According to the paper, the number of admissions for each day for which the principal diagnosis was a respiratory disease was abstracted from the central registry of information on discharges from hospitals kept by Statistics Canada.

Small positive associations were reported for Spring, Summer, and Fall but not for Winter. The association for April through December varied among cities ranging from 1.000 to 1.088 per 30 ppb increase in daily maximum 1-hour ozone. Associations were small or non-existent in Canada's second and third largest cities - Vancouver and Montreal. Positive associations were reported for 0, 1 and 2 day lags; associations with ozone were strongest with a lag of one day. Burnett et al. pointed out that ozone may be acting, in part, as a surrogate for other environmental factors that vary temporally. They also noted positive associations for several other pollutants with respiratory admissions in their study and in the general literature.

Evaluating the hospital admissions/ozone data base as a whole, the SP notes that the largest most significant associations for total respiratory admissions has been reported for 0 and 1-day lags. In many cases, positive associations are limited to the warm season. In

some studies, the positive association was limited to only a relatively small portion of the year, May through August in the case of Burnett et al. 2001. Focusing on only the months when there is a positive association gives a misleading picture of the consistency of the association, without a rational explanation for why ozone contributes to excess hospital admissions in some months and not in others.

There is insufficient discussion in the SP or the CD of the plausibility of 0 and 1-day lags for hospital admissions or other health endpoints. The key issue related to causality is whether the decision to be hospitalized occurs before the ozone exposure implicated in the association. A decision is made by a patient to initiate an emergency room visit. Some fraction of emergency room visits turn into hospital admissions, with the decision being made by emergency room staff. For other hospital admissions, the patient makes a decision to visit (or otherwise contact) his or her doctor and the doctor makes the decision at the visit to put the patient in the hospital. If there is a significant lag in either of these steps, the actual hospital admission may take place several days after the initiating event. Even in the case of a walk-in to an emergency room, if it occurs prior to the mid-afternoon ozone peak, a subsequent admission would not meet the temporality requirement.

Considerable information on the distribution of times for the various events involved in this sequence is needed to evaluate the plausibility of 0 and 1-day lags being causal. Fortunately, some of the required information is available. Stieb et al. 2000<sup>11</sup> have reported data characterizing cardiorespiratory disease episodes among patients visiting emergency rooms. They provide information on the interval between onset of symptoms and emergency room visit for several diagnosis groups. They included information on the earliest symptom and the most responsible symptom (the symptom that prompted the visit). For cardiac conditions the median interval between the earliest symptoms and visit was quite short, 0.8 days, but for respiratory infections or COPD the median interval was 4 days and for asthma it was 3 days. For COPD and respiratory infections, the median time interval between the symptom most responsible for the visit and the visit was 2 days.

With the addition of information on the distribution of time intervals between contacting the doctor and an office visit and between the time a doctor decides hospitalization is necessary and the actual admission, a more formal evaluation of the plausibility of 0 and 1-day lags for respiratory admissions could be carried out. In the meantime, the 0 day lag associations for hospital admissions appear problematic.

The review of ED studies during the 2004 ozone standard review in California indicated that if hospital admissions are affected by ozone, then it is likely that ED visits would also be affected, and to a greater degree. However, the ED results are less consistent, and summarized as being inconclusive, as noted in the SP. The California review noted that of 20 studies they evaluated, 16 reported at least one significant association involving ozone. Given that studies without effects are difficult to publish and that these studies often include multiple comparisons, it is not surprising that there would be a number of positive associations in the literature. However, the California review also acknowledged

---

<sup>11</sup> Stieb, et al., Canadian Journal of Public Health, **91**, 107-112 (2000)

that there is no clear pattern of association with specific respiratory disease outcomes. In the 27 studies daily emergency department visits included in Table AX7-3 of the CD, there is a very wide range of results with some multi-year studies reporting positive associations in only some years, with some ozone associations not robust to inclusion of other pollutants, with many single-pollutant associations being positive for only one of many lags or age groups. Therefore, it is hard to accept this data as evidence of effects.

### **Section 3.3.1.2 Effects on the respiratory system from short-term exposures**

This section includes evidence from both controlled studies and observational (epidemiologic) studies. The CD indicates that the most clear cut and compelling evidence for human health effects in 1996 was from the controlled human studies. (Page 8-11 of the CD) This is still true today. In controlled human exposures, the first effects on lung function occur at 0.50 ppm in sedentary individuals. (Page 3-9 of the SP) With increased levels of physical activity and breathing rate, the threshold for the first effects decreases. However, the thresholds for the first effects in controlled studies are still considerably above the vast bulk of personal exposures of the general population. During the 1996 review,<sup>12</sup> EPA staff concluded that the lowest range within which 1- to 3-hour exposures to ozone at heavy exertion produce group mean statistically significant lung function decrements is 0.12 to 0.16 ppm. They concluded that the lowest range for 6- to 8-hour exposures at moderate exertion is 0.08 to 0.12 ppm. For respiratory symptoms, the USEPA staff concluded that the lowest range for 1- to 3-hour exposures at heavy exertion is 0.16 to 0.18 ppm and for 6-to 8-hour exposures at moderate exertion, it is 0.08 to 0.12 ppm. For biochemical indicators of pulmonary inflammation, the lowest level for short-term exposures was noted as 0.20 ppm and the lowest level for 6- to 8-hour exposures was 0.08 to 0.10 ppm. The SP repeats the 1996 conclusions regarding these thresholds on pages 3-63 to 3-65. The 2004 analysis of human studies carried out by the state of California in the review of the State's ozone standard came to essentially the same conclusions. In addition, the CD and SP note that the gradation of adversity of respiratory effects determined in the 1996 review (Tables 3-2 and 3-3) still appears valid and reasonable. Thus, the findings from the controlled studies have augmented but not materially changed the understanding of ozone respiratory effects.

The new information from observational studies of acute respiratory effects is reviewed in the following sub-sections.

#### **Lung Function**

While small lung function effects have been reported in many observational studies, given the reflex mechanism underlying these effects, the important issue is the medical or clinical significance of these effects rather than their statistical significance.

Although there is now a more extensive set of studies that report lung function effects than in 1996, the data is not fully consistent. In particular, there are several studies in California that report little or no effects. The Delfino study of asthmatics in Alpine and

---

<sup>12</sup> See 1996 USEPA Staff Paper at pages 149-150.

the Linn study in Rubidoux, Upland and Torrance are noted in the CD and are of particular import given the fact that Southern California ozone levels are the highest in the nation. Potential reasons why decrements in the performance of lung function tests are less evident in the high ozone areas of California should be discussed and evaluated.

There is a particularly important study of asthmatic wheezy and healthy children in Southern California that is cited in the CD and relevant to the discussion in the SP. The Health Effects Institute study by Avol et al.<sup>13</sup> evaluated potential acute respiratory effects of ambient ozone in the spring and the summer. About 200 10 to 12 year old children from Lancaster, San Dimas, Upland, Mira Loma, Riverside and Lake Arrowhead were studied in the spring and late summer of 1994 with monitored hourly ozone levels exceeding the federal 1-hour standard in each season and with peak hourly levels up to 200 ppb in the summer period.

There was no discernible pattern of diurnal lung function changes by season, by ozone level, or by health group. While there were individual improvements and decrements in lung function observed, there was a general tendency for improvement and not reductions in lung function.

There was also no consistent or clear role of ozone in producing changes in reported symptoms or medication use. There were increased symptoms and medication use on high ozone spring days but not high ozone summer days. The spring results may have occurred because the subjects were more sensitive in the spring or, as the HEI Review Committee suggests, a result of enhanced response to allergens from springtime pollen blooms. The review committee also noted that there was an increased frequency of inhaler use on low ozone days in the summer but not on high ozone days.

The authors note that the failure to detect clear evidence of acute pulmonary effects in the three groups may be due to the fact that ozone levels were not sufficiently elevated during the study, although they exceeded 80 ppb during the study with peak hourly values up to 200 ppb. The authors note that the children in their study, on average, did not spend the amount of time outdoors nor did they spend the time exercising at the high levels used in the chamber studies.

### Symptoms and inflammation

The SP overstates the consistency of the database on symptoms and inflammation relative to the description in the draft CD which in turn overstates the consistency in the database itself. Even though the SP indicates that asthma panel studies as a group indicate a positive association between ozone and respiratory symptoms in asthmatics (page 3-12), the results for respiratory symptoms in single pollutant models in asthmatic subjects in the CD are mixed and inconsistent. The text of the CD indicates that the results for cough shown in Figure 7-5 suggest an association while the results for medication use shown in Figure 7-6 are described as being less consistent. The draft CD summary notes

---

<sup>13</sup> E. Avol, et al., Acute effects of ambient ozone on asthmatics, wheezy, and healthy children, Health Effects Institute Report Number 82, 1998.

in conclusion that “the various studies seem to indicate that ozone concentrations are associated with respiratory symptoms and increased medication use in asthmatics.” The SP should not interpret the data as stronger or more consistent than noted in the final CD.

The SP refers to a number of studies that report “fairly robust associations “with daily symptoms /asthma medication use, even after adjustment for co-pollutants while also noting that nine other studies (described as well-conducted) have not found these effects. (page 3-12) The SP and CD stress the multi-city study of Mortimer et al. 2002. as providing a sample of asthmatics that is most representative of the U. S. The single pollutant model results from the Mortimer et al. study are included in Figure 3-3 of the SP. The CD text acknowledges that other pollutants were associated with symptoms (page 7-45) in the Mortimer et al. study but indicates that, in multi-pollutant models, the ozone effect was shown to be slightly diminished.

This last statement is wrong and highly misleading. Table 4 of the Mortimer study presents the results of single and multi-pollutant models. Because the various possible confounders were not all measured in the eight cities, the results are shown for various subsets of the data. Only SO<sub>2</sub> was measured in all eight cities, and when SO<sub>2</sub> was included in two pollutant models, the ozone coefficient was reduced from 1.16 to 1.11 and became non-significant. For seven areas both SO<sub>2</sub> and NO<sub>2</sub> were available. In multi-pollutant models the ozone coefficient was reduced to 1.06 and again was non-significant. In three cities PM10 was also available, and in those areas the ozone coefficient was reduced to 1.04 in a two-pollutant model with PM10 and to 1.00 in a multi-pollutant model with all 4 pollutants. As the authors indicate, only SO<sub>2</sub> remained significant in the seven urban areas with complete data for the three gases. Overall, each of the four pollutants was associated with symptoms in single pollutant models and the strongest association was with NO<sub>2</sub> in single-pollutant models and with SO<sub>2</sub> in multi-pollutant models. Thus, the CD’s characterization of the influence of ozone versus other pollutants on symptoms is misleading. In fact, the authors of the study specifically mention that they focus on ozone as a good marker of summer air pollution and refer to the study’s finding as effects of summer-time air pollution rather than ozone, per se.

Because of the over-emphasis on the Mortimer study of symptoms and the listing of nine other studies that are described as having limited or a lack of evidence for ozone effects, the overall findings relative to effects of ozone on asthmatics on symptoms in the SP should be described as mixed and inconsistent. Confounding by bioaerosols may a possible explanation for some of the variation in results for asthmatics. Some bioaerosols are known to exacerbate asthma but they are generally not considered or included in air pollution studies. Based on studies with healthy children, the CD correctly notes that there is no consistent evidence of an association between ozone and respiratory symptoms in healthy children. The SP should include this conclusion.

While there is ample evidence of inflammatory responses resulting from high dose exposures of exercising subjects in controlled studies, the few observational studies available on nasal or other inflammatory endpoints that might be associated with ozone do not support the view that ozone is causing increased inflammation in the population on

a day-to-day basis. In addition, as Mudway and Kelly 2004 have shown, the inflammatory responses are a threshold phenomenon.

#### Section 3.3.1.2.7 Increased school absences

The text on page 3-18 indicates that three studies suggest that ozone, on the same day or accumulated over 2 to 4 weeks, is associated with an increase in school absences, particularly for respiratory illnesses. The text further notes that further replication is needed before firm conclusions can be drawn.

The first study discussed by the text is that of Chen et al. 2000. This study is weak because it does not separate out absences due to illness from all absences. The ozone associated reported was for a distributed lag of 1-14 days, which is problematical. In addition, associations were reported for several other air pollutants.

The Gilliland et al. 2001 study of school absences is discussed next. It is noted in the CD as being especially valuable. However, as acknowledged in the CD, there is a question of residual seasonal confounding for this study. In addition, the CD indicates that the biological relevance of ozone lagged thirty days presents a challenge. Furthermore, the reporting of a substantially larger effect in low ozone communities versus the effect in high ozone communities, a highly counterintuitive result, necessitates additional explanation. Potential confounding by bioaerosols is mentioned in the CD but was not evaluated in the study. Thus, it would be particularly important to replicate the study in another cohort and over all seasons with additional potential causal covariates. The many issues with model selection that have arisen in recent years raise additional concerns that the result may not be robust.

In fact, there is now direct evidence that the finding is not robust. Two additional studies report analyses of the same Children's Health Study data. Berhane and Thomas<sup>14</sup> reported a nearly significant acute ozone association with a 30-day distributed lag ( $p = 0.075$ ) but no chronic (between community) effect. Rondeau et al.<sup>15</sup> reported no acute ozone association with a 30-day lag but did report a chronic (between community) effect. In the Berhane and Thomas analysis, the acute association with a 15-day lag was not significant and even in the 30-day analyses, the largest and only statistically significant association was in the lowest pollution (low ozone and low PM/NO<sub>2</sub>) communities. In addition, Berhane and Thomas report the individual-community distributed-lag effects which show a mixture of highly concave and highly convex shapes. While the individual-community results are a mix of strong positive and strong negative associations for lags 1, 2, and 3, the combined association is actually negative at all three lags. Such a wide variation in associations and an overall negative association for the time period where an absence effect might logically be expected is not supportive of a causal ozone effect on school absences.

---

<sup>14</sup> K. Berhane and D. Thomas, *Biostatistics*, **3**, pages 21-32 (2002)

<sup>15</sup> V. Rondeau, K. Berhane, and D. Thomas, *Statistics in Medicine*, **24**, pages 1103-1115 (2005)

Rondeau et al. evaluated 5-day, 15-day and 30-day distributed lags and report that they found no associations of acute ozone with school absences, either total absences, illness-related absences or respiratory-illness related absences. They note that their results differ from those of Gilliland et al. and Berhane and Thomas and discuss the possibility that the different modeling approaches are responsible for the different results. In particular, they mention that Berhane and Thomas allowed the distributed lag parameters to vary by community while they did not. Since widely different response patterns are not biologically plausible between near-by communities, the Rondeau approach would seem to be preferred.

There are substantial other differences in the modeling approaches among the three analyses. They all are attempts to use state-of-the-art sophisticated statistical analysis techniques to analyze observational data. Yet, the results differ dramatically. This is an excellent example of the model selection problem that plagues the field of air pollution epidemiology. All three studies should be discussed in the CD and in the SP.

The large reported ozone effect on respiratory absences in Gilliland is disproportional to the lack of substantial effects on respiratory symptoms seen in school children residing in the Basin. Chapter 7 of the CD acknowledges that there are several studies of respiratory symptoms in California that failed to detect symptomatic effects. These include Delfino et al. 1996, Delfino et al. 1997b, Linn et al. 1996, and Ostro et al. 2001. There is a particularly important study of asthmatic, wheezy, and healthy children in Southern California that is not sufficiently discussed in Chapter 7. The Health Effects Institute study by Avol et al.<sup>16</sup> evaluated potential acute respiratory effects of ambient ozone in the spring and the summer. About 200 10 to 12 year old children from Lancaster, San Dimas, Upland, Mira Loma, Riverside and Lake Arrowhead were studied in the spring and late summer of 1994 with monitored hourly ozone levels exceeding the federal 1-hour standard in each season and with peak hourly levels up to 200 ppb in the summer period.

There was no consistent or clear role of ozone in producing changes in reported symptoms or medication use. There was also no discernible pattern of diurnal lung function changes by season, by ozone level, or by health group. The authors note that failure to detect clear evidence of acute pulmonary effects in the three groups may be due to insufficient elevation of ozone levels during the study, although ozone concentrations exceeded 80 ppb during the study. The authors note that children in their study on average did not spend a comparable amount of time outdoors or exercising at the high levels to the time used in the chamber studies.

The lack of a substantial respiratory symptom signal in several Southern California studies is inconsistent with the large effect on school absences reported by Gilliland. The varying results of the three studies that analyzed the same data set are also troubling. These issues need to be acknowledged and discussed in the SP.

---

<sup>16</sup> E. Avol, et al., Acute effects of ambient ozone on asthmatics, wheezy, and healthy children, Health Effects Institute Report Number 82, 1998.

The third study noted is Park et al. 2002. This study should be discounted because it reports the strongest association with same day ozone. However, for school absences that are determined in the early morning, an association with afternoon ozone violates the temporality requirement. Thus, the reported association cannot be causal.

When all five studies are considered along with other air pollution/school absence studies in the literature, referenced in Park et al. 2002 and indicated as producing inconsistent results, the case for an ozone/school absence association is much weaker than the current SP indicates.

### **Section 3.3.1.3 Effects on the respiratory system from long-term exposure**

Because of the presence of temporary decrements in the performance of lung function tests in both the controlled studies and evidence of small changes in the observational studies, the question of whether there are permanent changes in lung function caused by ozone is of paramount importance. The text on page 3-20 correctly concludes that there is little evidence for long-term lung function decrements. Three Gauderman et al. publications from the Southern California Children's Health Study clearly demonstrate the lack of an effect in a well-designed study.

The text should also inform the reader that there is evidence for possible effects of other pollutants on long-term development of lung function in children. This is important because of the issue raised by Tager that it is not clear whether a published ozone association reported by Frischer et al. 1999 with lung function changes represents ozone alone independent of other pollutants or not. The Gauderman et al. Southern California studies by including other pollutants specifically addressed this question. It is also important to put the ambient ozone levels measured in the California communities into perspective in terms of whether the current 8-hour standard was exceeded in these communities and by how much.

Finally, possible reasons why no long-term lung function decrements have been seen should be addressed in the CD. The likelihood that the reflex neural mechanism is responsible for the shorter-term effects and that mechanism would not be expected to cause long-term effects should be mentioned.

### **Section 3.3.1.4 Effects on the cardiovascular system**

As noted above, several CASAC panelists indicated during the December 6-8, 2005 meeting that the second draft CD and the first draft SP over-interpreted and overstated the case for cardiovascular effects. The limited database reporting small changes in cardiac parameters associated with ozone or other pollutants in observational studies could easily be confounded by a host of factors, and are not found consistently for any specific pollutant. In addition, the changes in heart rate or various measures of heart rate variability are small and within normal background noise due to changes in day-to-day activities and are too small to be of clinical significance.

For cardiovascular hospital admissions, the range of associations from time-series studies is very large, from -12 to + 35 % per standard ozone increment as shown in Figure 7-13 of the CD. The data from ten studies referred to in the text of the CD are described as inconsistent or negative. Other studies have reported strong positive associations. There is no discussion of potential confounding or model selection issues in these studies. Overall the evidence is quite inconsistent (as noted in the SP at page 3-24) and not suggestive of a causal association.

As we noted above, the controlled exposure study by Gong et al. 1998 does not suggest any effects from typical day-to-day ozone exposures. Gong et al.<sup>17</sup> exposed subjects to 0.3 ppm ozone for 3 hours with exercise and found little direct effect of ozone on the heart or vasculature for hypertensive or healthy subjects even though many pertinent variables (such as cardiac index, ventricular performance, pulmonary artery pressure, pulmonary and systemic vascular resistances, ECG, serum cardiac enzymes, plasma catecholamines, and atrial natriuretic factor, for example) were evaluated. Gong concluded that they did not find convincing evidence of significant direct cardiovascular effects from ozone exposure, analogous to effects reported in laboratory animals. The small indirect effect they report, impaired gas exchange, can be explained as an effect secondary to the respiratory effects of the high ozone exposures.

The extremely low personal exposures of the population on a day-to-day basis are not expected to cause any cardiovascular effects. The text in Section 3.3.1.4 should be changed to acknowledge the major inconsistencies in the observational studies and the lack of effects in controlled human studies.

### **Section 3.3.2 Premature mortality**

#### **Multi-city analyses**

The SP places significant emphasis on the question of ozone mortality effects, especially from the results of multi-city analyses. It refers to the original NMMAPS analysis, the NMMAPS re-analysis, a new NMMAPS analysis<sup>18</sup> that extends the previous analyses, and analysis of a subset of 19 large U. S. cities (Huang et al. 2005). AIR has evaluated these various analyses and would like to point out a number of issues and concerns with interpreting the results from them as causal.

The Bell et al. 2004 paper is the third major analysis of ozone and mortality with the NMMAPS database. The first two analyses were sponsored by the Health Effects Institute. The first NMMAPS study was carried out with the General Additive Model (GAM) that was subsequently identified as having statistical problems that led to a re-analysis of selected studies. In the GAM study, selected ozone results were presented. For lag 0 with 90 city data, there was no overall ozone/mortality association although

---

<sup>17</sup> A full report of the Gong et al. 1998 study is available as a report to the California Air Resources Board, Final Report Contract No. 93-327, May 2000.

<sup>18</sup> M. L. Bell, et al., "Ozone and Short-term Mortality in 95 US Urban Communities, 1987-2000," JAMA, **292**, 2372-2378, 2004.

there were positive associations for PM<sub>10</sub>, CO, NO<sub>2</sub>, and SO<sub>2</sub>. (Figure 9 of NMMAPSII<sup>19</sup>). For the 20 largest cities, ozone was reported to have a positive combined association in summer at lag 0 but a negative combined association in winter, with a small negative association in year round data. (Figure 30 of NMMAPSII) For lags 0,1, and 2 in the 20 cities, ozone had a small negative association in year round data at all three lags. (Figure 26 of NMMAPSII). In addition, in a paper<sup>20</sup> presenting the NMMAPS results for the 20 largest cities, the single pollutant model ozone association was negative in the six California cities included in the analysis. Thus, there was little evidence of an ozone effect except for the summer results, which were balanced by negative associations in winter analyses.

In the second study, the HEI re-analysis<sup>21</sup> using the Generalized Linear Model (GLM), the investigators reported a negative combined association of ozone with mortality in the winter for lag 0 and a positive combined association in the summer, but with a small positive association for lag 0 when all seasons were included. However, in multi-pollutant models, the combined ozone effect was reduced and became non-significant for lags 0, 1, and 2 for the summer data. (see Figure 7-28 of the CD) Results for winter or all year in multi-pollutant models were not reported. Overall, the NMMAPS analyses were focused on PM<sub>10</sub> with limited analyses reported for ozone or other pollutants. The re-analysis reduced the overall PM<sub>10</sub> association by about 40 % but it increased the overall ozone association from slightly negative to slightly positive. In both GAM and GLM analyses, a distinct difference between summer and winter was reported with ozone having a positive association with mortality in summer but negative in winter.

The Bell et al. 2004 paper extended the ozone mortality studies in several ways but also omitted winter ozone results and limited the analyses regarding other pollutants. While the additional analyses are welcome and are discussed below, the omission of winter and comprehensive multi-pollutant results is a serious deficiency in the study. It is difficult to interpret the overall positive associations in summer as causal when there are overall negative associations of ozone with mortality in winter. It is also difficult to interpret the positive ozone associations in summer as causal when the combined association is reduced and non-significant when other pollutants are added into the mix as reported in the second HEI analysis. Although Bell et al. 2004 is an extension of the HEI NMMAPS ozone work, the paper neglects to mention that ozone had a combined negative association with mortality in winter, as reported in both the original GAM and the re-analyzed GLM work.

Whereas the HEI re-analysis reported ozone associations for single day lags 0, 1, and 2, Bell et al. reports results for these three lags plus lag 3 and a distributed lag model that includes the same day and the 6 previous days. As shown in their Figure 1, Bell et al. found the strongest association for the same day (lag 0) with progressively smaller

<sup>19</sup> J. M. Samet, et al., The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity and Mortality from Air Pollution in the United States, Health Effects Institute Research Report 94, Cambridge, MA, 2000.

<sup>20</sup> M. Daniel, et al., Am. J. Epidemiology, **152**, 397-406, 2000

<sup>21</sup> HEI Special Report, Revised Analyses of Time-Series Studies of Air Pollution and Health, Health Effects Institute, Cambridge, MA, May 2003, page 18.

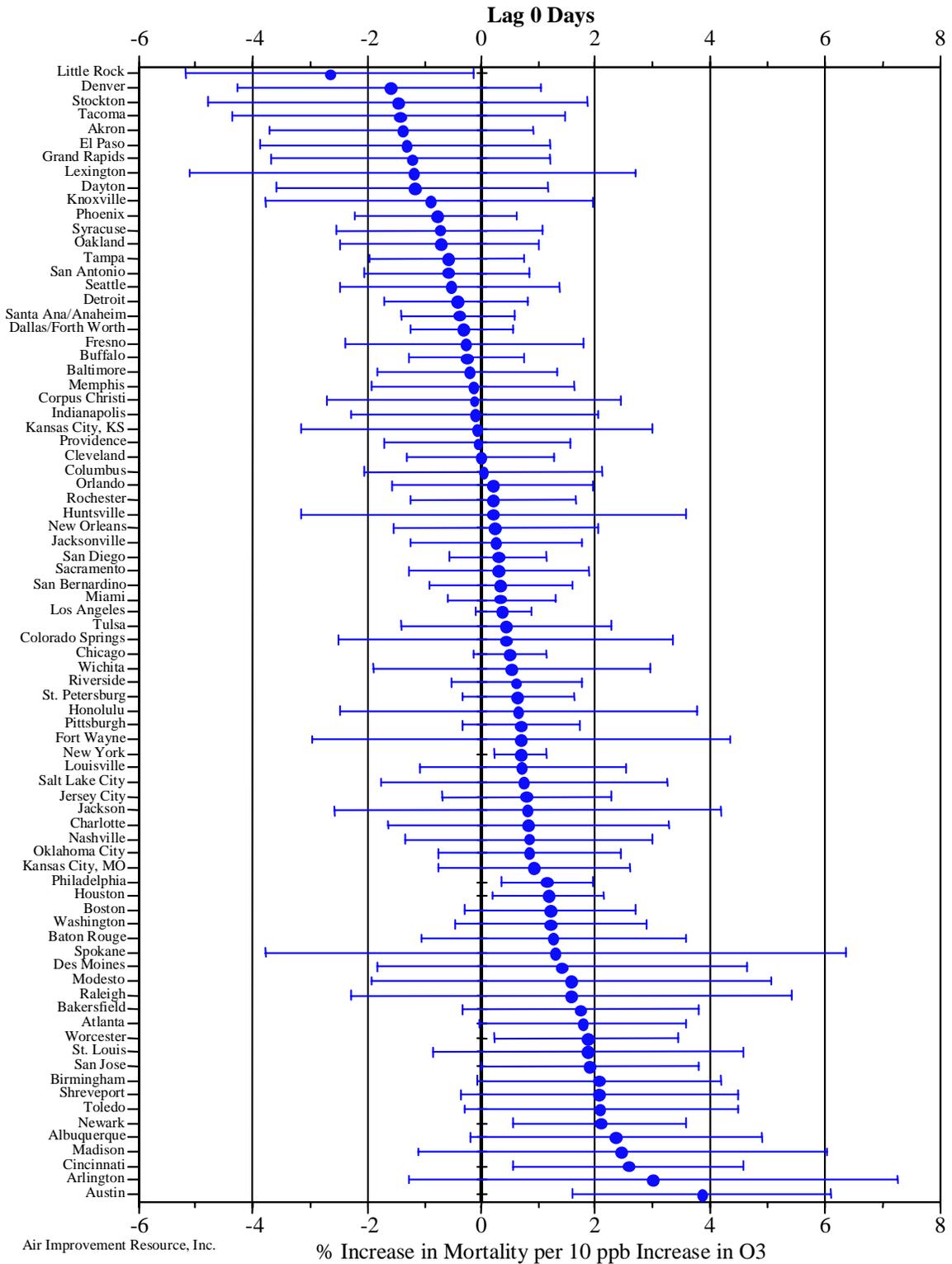
associations through lag 3. Beyond lag 2 the individual day associations were not significant. The effect from the distributed lag model for a 10 ppb increase in ozone over the previous week is just slightly higher than the effect for a 10 ppb increase on lag 0 and lag 1. Thus, the Bell et al. analysis which considered more ozone lags is consistent with the HEI analysis indicating the strongest association is on day zero. This is problematic on two counts. First, since ozone peaks in the afternoon, a same day association risks violating the temporality requirement (that the exposure preceded the response). Second, if there is a causal link between a 10 ppb increase in ozone on the same day and premature death, it implies that ozone exposures that are well within background levels are causing or contributing to death immediately. There is nothing in the clinical or animal toxicology literature that suggests this could be happening.

Bell et al. presented individual city results only from a second stage analysis. This is reproduced as Figure 7-16 in the CD. In the first stage analysis, however, there is a biologically implausible large range (from negative to positive) in individual city results. The individual city results for the summer in the HEI NMMAPS re-analysis were posted on the Johns Hopkins website for lags 0, 1, and 2. As shown in those data, reproduced below as Figures 9 to 11, the pattern of results for all three lags is the same, with a range of ozone associations from about 3 % negative per 10 ppb increase in daily ozone to 3 % positive association with a 10 ppb increase in daily ozone. In terms of the standard increments used in the CD, the range for each day is from about - 6 % to + 6%. The interpretation of the positive associations in some cities as adverse health effects would have to be balanced by an interpretation of negative associations as health benefits. Since health benefits from ozone exposure are implausible and the overall wide range of results is equally implausible, interpreting these data as a causal effect of ozone is problematic.

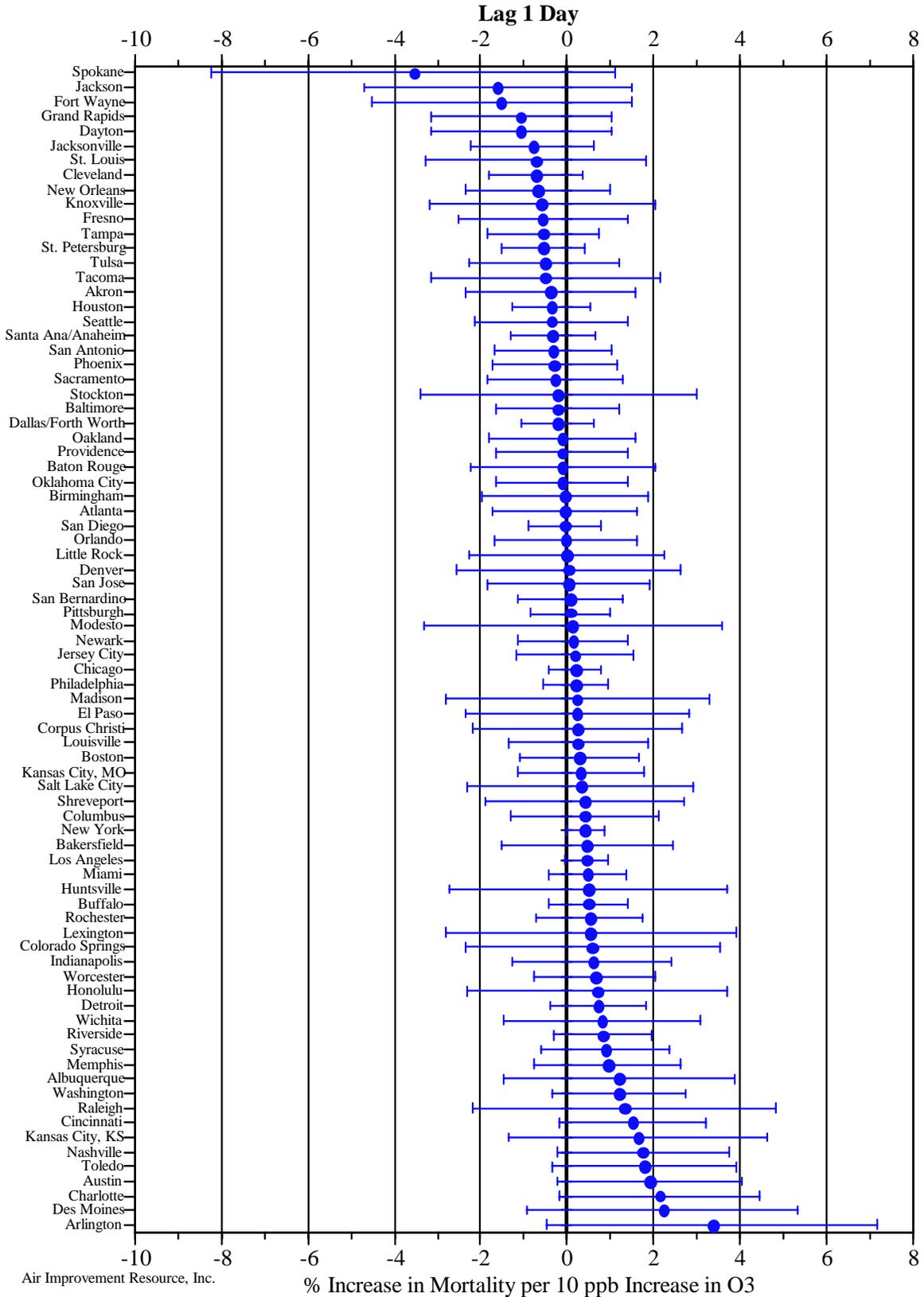
The results from the HEI NMMAPS re-analysis for lag 0, the day with the largest effect, and lags 1 and 2, which are more typical, are reproduced below in Figures 9 to 11. Note that only 7 of the 80 cities have statistically significant positive associations with ozone for lag 0 in the summer. The Bell et al. paper masks the wide range in individual city results by using a second stage Bayesian analysis to derive the results presented in their Figure 2 (Figure 7-16 of the CD). The existence of a biologically implausible wide range of results from individual cities is shown, however, in their Figure 3 (Figure 7-29 in the CD) in which effects from - 5 % per 10 ppb increase to + 10 % per 10 ppb increase are reported. Even in the second stage analysis shown in their Figure 2, only 6 of the individual cities have statistically significant associations in single pollutant models.

**Figures 9-11**

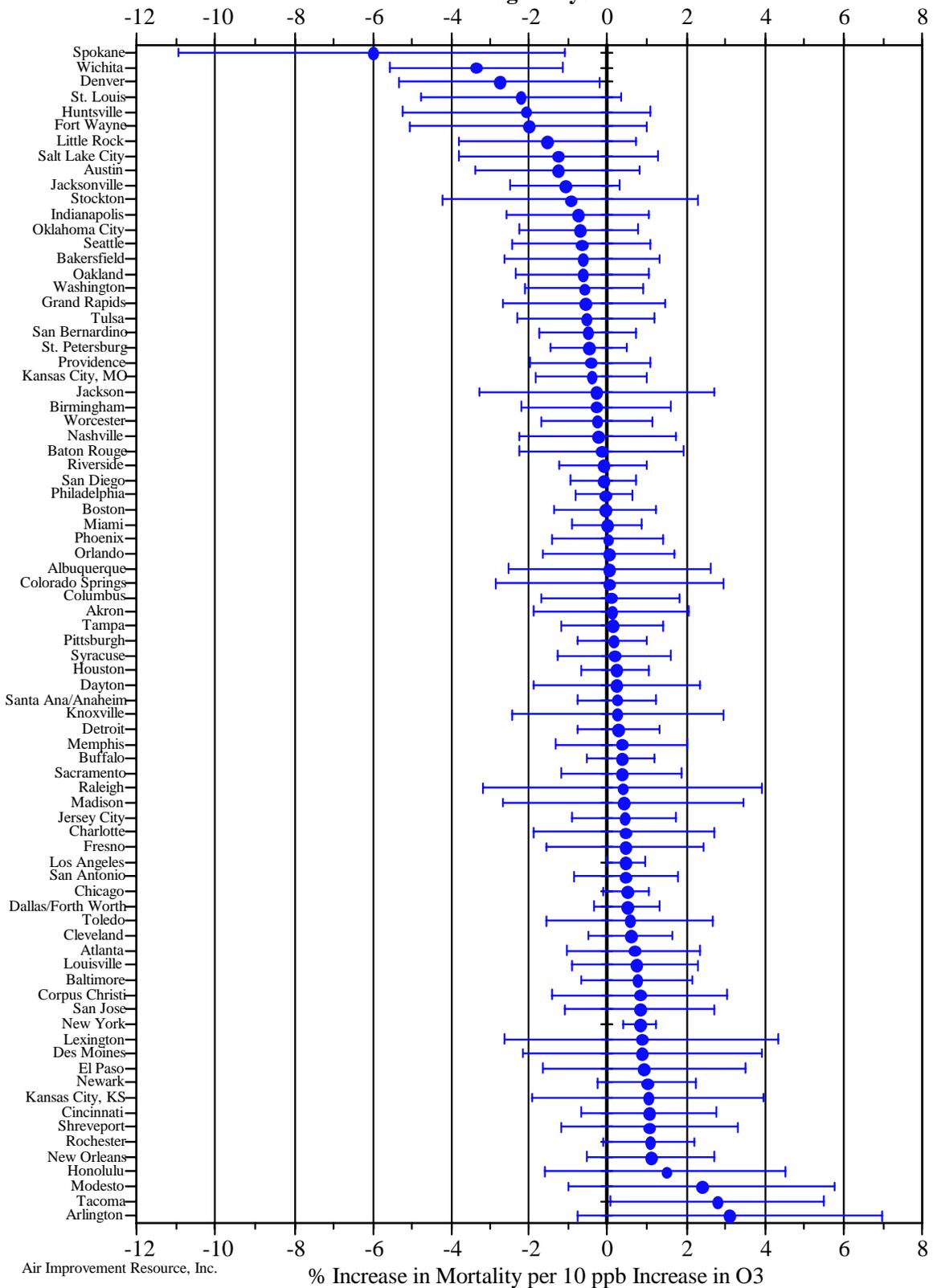
**NMMAPS Maximum Likelihood Estimates and 95% Confidence Intervals of the Percentage Increase in Total Mortality from Nonexternal Causes per 10 ppb Increase in Ozone Concentration for Each Location**



**NMMAPS Maximum Likelihood Estimates and 95% Confidence Intervals of the Percentage Increase in Total Mortality from Nonexternal Causes per 10 ppb Increase in Ozone Concentration for Each Location**



**NMMAPS Maximum Likelihood Estimates and 95% Confidence Intervals of the Percentage Increase in Total Mortality from Nonexternal Causes per 10 ppb Increase in Ozone Concentration for Each Location Lag 2 Days**



Bell et al. report that their combined result was robust to exclusion of high temperature days, to exclusion of high ozone days, to the use of different ozone metrics from daily averages to 1-hour maxima, and to inclusion of PM10 in the model. The combined ozone association was similar for total deaths and for cardiovascular plus respiratory deaths. The association was similar for all three age groups evaluated. If this association is a real health effect, it implies that ozone is contributing to cancer and diabetes deaths as much as cardiovascular and respiratory deaths and to deaths in the younger and the aged population equally. Yet given the negative association in winter, somehow the ozone effect is strongly seasonal, being beneficial in winter. Given what we know of air pollution health effects, this combination of results is not plausible. The combined effect probably represents residual confounding in a noisy ecologic analysis. For example, Lumley and Sheppard<sup>22</sup> caution that “estimation of very weak associations in the presence of measurement error and strong confounding is inherently challenging. Prudent epidemiologists should recognize that residual bias can dominate their results.”

If this reported association is causal, it likewise implies that ozone levels well within background are linked to premature mortality. Since the ozone exposures of the frail population (since they spend their time indoors) will be roughly half or less than that measured at central monitors, the paper implies that levels well below background (average exposures of 10 to 15 ppb or 0.01 to 0.015 ppm) are causing significant mortality. This is not biologically plausible. Bell et al. argue that the pattern of association they observe would be anticipated for ozone since ozone produces acute inflammatory responses in the lung. However, the doses of ozone that result in initial inflammatory changes are considerably elevated compared to the doses the frail population experiences indoors on a day-to-day basis. As noted in the 1996 Staff Paper, 0.20 ppm is the lowest level at which 1- to 3-hour exposures to ozone of subjects engaged in very heavy exertion have induced biochemical indicators of pulmonary inflammation. For 6- to 8-hour exposures of subjects engaged in moderate exertion, the range of concentrations that induced this effect is 0.08 to 0.10 ppm. Without heavy or sustained exercise, the threshold for these physiological changes is substantially higher. In evaluating the medical significance of these first inflammatory changes, EPA indicated in the 1996 Staff Paper that their view, along with that of medical experts they consulted, was that acute inflammatory responses, only when repeated, were a potential public health concern.

In contrast to the HEI NMMAPS analyses, which evaluated models including other gases as potential confounders, the Bell et al. paper only looked at PM10 as a potential confounder. Since there is greater concern over fine PM than coarse PM as causing premature mortality, a test for confounding by PM10 is not an adequate test even for just PM. Both earlier HEI NMMAPS analyses evaluated a wider range of air pollutants as confounders. As shown in Figure 7-28 of the CD, none of the combined ozone effects in various multi-pollutant models were statistically significant. In addition, in the Stieb et

---

<sup>22</sup> T. Lumley and L. Sheppard, *Epidemiology*, **14**, 13-14, 2003.

al.<sup>23</sup> meta-analysis, a comparison of ozone single-pollutant models with multi-pollutant models including ozone was reported. The combined ozone association from single pollutant models was reduced and became non-significant in multi-pollutant models.

Bell et al. acknowledge that the ozone association they report may not be from ozone per se but may represent the risk from the photochemical pollution mixture more generally. This interpretation is also subject to criticism. It is one possible explanation for the combined negative association in winter but it does not explain the implausibly wide range of negative and positive associations among cities in all seasons. Bell et al. also raise the possibility that the ozone signal may be a surrogate for other pollutants or the pollution mixture in general. For all the reason enumerated above, it is premature to interpret the ozone associations reported by Bell et al. as causal.

The SP quotes the combined summer ozone associations with cardiovascular plus respiratory mortality from the Huang et al. study of 19 large cities. However, the pattern of individual-city associations in the Huang study is not consistent with a causal association. For example, the range in individual-city associations is from - 3 % to + 8 % and only 5 of the 19 studies had statistically significant positive associations. Four cities had negative associations. Most importantly, the three strongest associations (each around 8 %) were in Oakland, Seattle, and San Jose, three cities with among the lowest ozone levels in the study (see Figure 1 of Huang et al.). In contrast, the city with the highest ambient ozone, San Bernardino, had an ozone association of - 1 %.

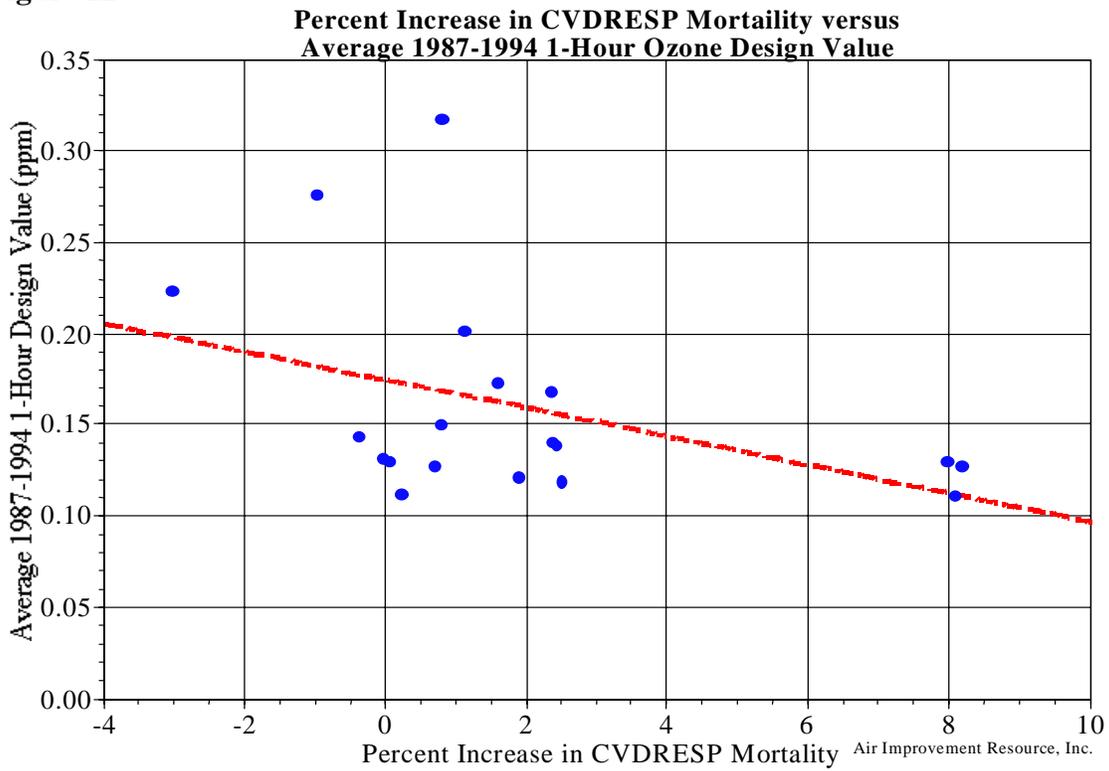
To further explore what appears to be a pattern of stronger associations in cities with lower ozone, we plotted the cardiorespiratory mortality associations from Table 1 of Huang et al. versus the 1-hour and 8-hour design values for the areas included in the study averaged over the period of the study. The results are shown below in Figures 12 and 13. In each case there is a negative slope to the linear regression through the data indicating that lower ozone is associated with higher mortality associations. As noted above, such a pattern is not consistent with the associations being causal.

The SP also discusses the Gryparis et al. and Schwartz multi-city studies. Both studies also reported a very wide range of individual city associations ranging from negative to positive. For example, the percentage increase in deaths for a 40 ppb increment in 1-hour ozone ranged from - 12 % to + 6 % in the year round data as reported in 23 European cities by Gryparis et al. (see their Figure 1). For the summer period, the range was from - 8 % to + 8 % (see their Figure 2). In the Schwartz 2004 14-city study there was also a wide range in associations ranging from positive to negative (see his Figures 1 and 2). Schwartz notes the negative association in winter and suggests that it may reflect the negative association between wintertime ozone and primary air pollutants. He goes on then to ask “might not the positive association in the summer likewise reflect confounding with some other pollutant?” These and other possible explanations need to be rigorously considered in the SP.

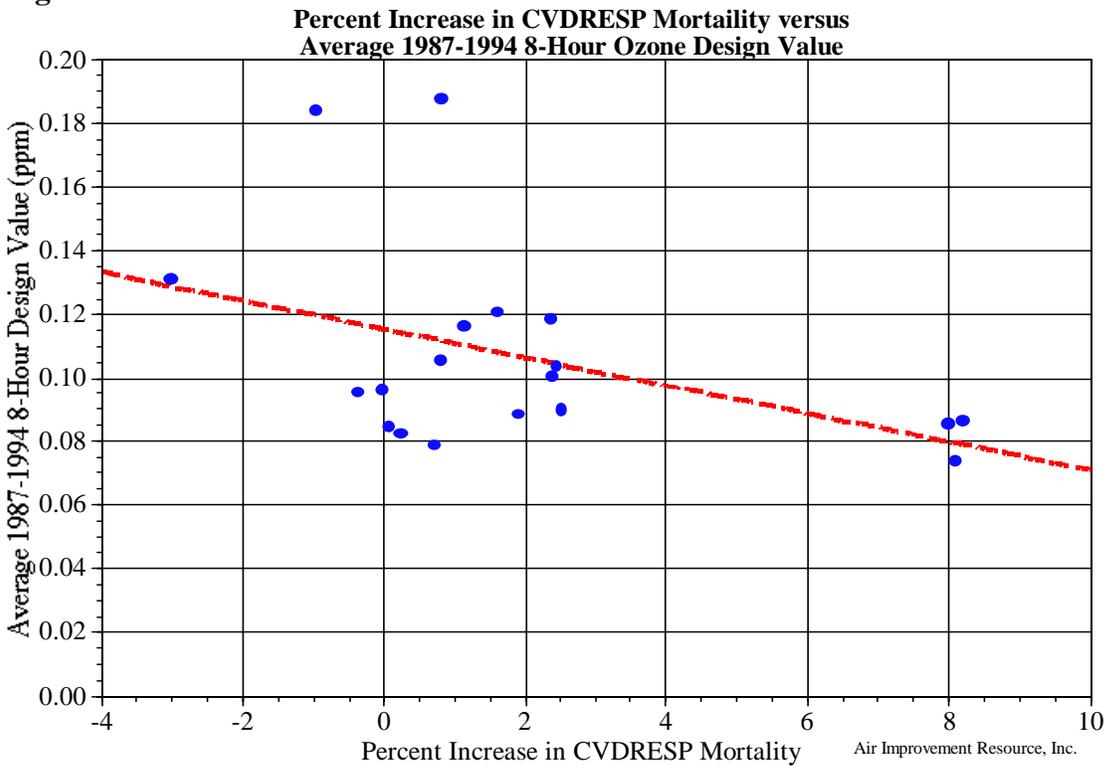
---

<sup>23</sup> Stieb et al., *J. Air & Waste Management Association*, **52**, 470-484, 2002; Stieb et al., *J. Air & Waste Management Association*, **53**, 258-261, 2003.

**Figure 12**



**Figure 13**



## Meta-analyses

The SP also discusses three new meta-analyses and points out that they come to similar conclusions. When these studies were published they were accompanied by two commentaries that are referenced in the CD but not the SP. The commentary by Goodman is particularly insightful. It notes that the implications of the EPA-sponsored exercise of funding three separate meta-analyses “go far beyond the question of the ozone mortality effect.” He cautions that “depending on published single-estimate, single-site analyses is an invitation to bias.” He notes that “the most plausible explanation is the one suggested by the authors, that investigators tend to report, if not believe, the analysis that produces the strongest signal; and in each single-site analysis, there are innumerable model choices that affect the estimated strength of that signal.” He goes on to note that both Bell et al. and Ito et al. provide empiric evidence within their meta-analyses to support this explanation. Although the SP mentions the issue of publication bias in this section it is discounted by noting that ozone-mortality associations remained after accounting for the bias.

However, the issue cannot be dismissed so easily. In addition to Goodman, Levy et al. 2001<sup>24</sup> raise the issue of publication bias. Levy et al. describe their review of the ozone health effects literature as providing weak epidemiologic support for ozone mortality and as potentially biased because it excludes studies that found ozone to be insignificant. After listing a number of such studies and noting that nearly all the studies found ozone not to be a predictor of mortality including all studies in warmer climates, Levy et al. concludes this information must “...give one pause in using our pooled estimate as a representative measure of ozone mortality risks.” They also discuss the difficulties of separating temperature or weather and ozone effects, cautioning that “possible confounding by weather leaves lingering doubt about whether the ozone findings are causal, an issue that would be difficult to resolve epidemiologically.”

In addition, a World Health Organization meta-analysis cited in the CD discusses the issue of publication bias acknowledging that it is a common and possibly universal problem in our research culture. Furthermore, they note that there are particular reasons why it might occur in time series studies. These include the large number of associations evaluated and the possibility for a selection bias involving the results included in a paper. They also mention a lag selection bias that can be overcome with multi-city studies such as NMMAPS. They further point out that publication bias can result in false conclusions being drawn in the hazard identification stage of a risk assessment and result in an inflation of the magnitude of the health impacts. They also recognize that the methods of detecting and correcting for publication bias are not without problems.

Since publication bias is a major factor overstating the magnitude of air pollution health effects in epidemiologic studies, the strength of associations reported from meta-analyses should be severely discounted. In addition, the same tendencies involved in fitting individual-city results to find the strongest signal can also be involved with multi-city

---

<sup>24</sup> J. I. Levy, et al., *Environmental Health Perspectives*, **109**, 1215-1226 (2001).

studies. So although publication bias is not a concern with multi-city studies, a model selection bias is still a concern.

### Summary of acute mortality analyses

The miss-characterization of the consistency of the associations needs to be corrected. The summary of this section (page 3-28) indicates that the epidemiologic evidence shows robust associations with daily ozone concentrations and mortality, with effect estimates ranging from 0.5 to 2.5 % in multi-city studies and from 0.5 to 5 % in individual-city studies. This is a gross miss-characterization of the data in the literature. For individual-city associations, the range is – 8 to + 7 % as shown in Figure 7-14 of the CD. For multi-city studies, the range of individual-city results is from – 12 to + 8 %.

### Section 3.3.3 Summary of effects

The summary on pages 3-30 and 3-31 takes the overstatements from the individual sections and subtly changes the wording to make even more of an overstatement of the evidence for ozone health effects in the scientific literature. For example, the inconclusive evidence for respiratory ER visits becomes “positive but less conclusive than the evidence for hospital admissions.” The discussion of acute mortality refers primarily to the meta-analyses without noting that they are subject to major publication bias. The conclusion that consistent associations have not been reported between long-term ozone exposure and mortality on page 3-30 becomes “the limited evidence from long-term studies... does not provide conclusive evidence of an association with mortality.” All these overstatements of the magnitude and certainty with which the health effects of ozone have been established should be removed from the SP.

At the December 6-8, 2005 CASAC meeting, the panel urged EPA to insure that the scientific findings from the body of the CD were summarized accurately and consistently as the material went from the detailed discussion in the annexes to the more limited discussion in the main chapters and then to the Executive Summary. The Agency must also insure that the material is accurately portrayed in the SP both in the main chapters and the summary sections. We have provided numerous examples of instances where important facts or important qualifiers are omitted as the material in the CD is discussed and summarized in the SP. The overall result is that the current draft SP significantly overstates the strength and consistency of the epidemiologic database concerning potential ozone health effects.

### **Section 3.4 Integrative assessment of evidence from epidemiological studies**

As documented in the following sections, the same tendency to overstate the case for ozone health effects by omitting facts or omitting important qualifiers carries over to Section 3.4.

### Strength of association

The discussion of the strength of associations in Section 3.4.1 quotes a range of mortality associations from 0.5 to 5 % when the full range in individual-city associations in the literature cited in the CD is from – 12 % to + 17 %. As we have shown above, the full range of associations for each health endpoint is very large ranging from negative to positive. In addition, the influence of publication bias is ignored.

### Exposure error

In the discussion of exposure error (Section 3.4.2.1) the main point made is that ambient ozone concentrations “may serve as a valid surrogate” for use in time-series studies but, since ambient concentrations generally overestimate true personal exposures, the time-series risk estimates may be underestimating true public health risk. This conclusion overstates the case for use of ambient measurements in time-series analyses and presents an incomplete discussion of the ramifications of the ambient ozone/personal ozone relationship.

There are two major questions involved in the use of central monitoring data to estimate exposure. One is whether the central monitoring data is a useful surrogate for use in epidemiologic studies and the second is how the actual personal human exposures compare to the central monitoring data. During the PM review, an entire chapter of the PM CD was devoted to these questions. In the ozone SP there is insufficient consideration of these questions. While there have been a variety of studies of how ozone at central monitors compares to personal exposures with a variety of different results, the text in the CD provides only superficial summaries of the studies and no data is shown. In some cases, correlation coefficients are noted but in other cases the correlation is only discussed using very general qualitative terms. In the end, the text indicates that while there are many issues related to using central monitoring data, and ozone/exposure may differ from the concentrations measured at central monitors, until more research is conducted, the use of monitored ambient concentrations as a surrogate for exposures is not expected to change the principal conclusions from ozone epidemiologic studies. (page 3-76 of the CD)

Several relevant issues are discussed in the CD. They range from the general comment that there is no clear consensus as to how well stationary monitor measurements represent a surrogate for personal exposures (page 3-69) to the more specific comment that, because of spatial variability, caution should be exercised in using data from a network of monitors to approximate human exposure (page 3-14). Importantly, page 7-8 of the CD includes the statement that “there is suggestive evidence” that ambient ozone may serve as a valid surrogate for personal exposures in time-series studies. When translated to this section of the SP, all these qualifying comments disappear and the conclusion is drawn that ambient ozone may serve as a valid surrogate. As noted above, this dramatically overstates the case. In addition, elsewhere in the SP, staff indicates that “personal exposure measurements of ozone are often not well-correlated with ambient measurements.” (page 4-14)

Another issue that is clearly outlined in the CD but not mentioned in the SP is that ozone is often measured at heights between 3 and 15 meters and people's exposure outdoors occurs at 1-2 meters. (see page 3-15 and page AX3-190 of the CD) This is particularly relevant to both the discussion and interpretation of the epidemiologic studies and the calculation of human exposures in Chapter 4 of the SP. Since the CD raises serious questions concerning the use of central monitor data as a surrogate for personal exposures, the SP should not just take one quote from the CD that implies that there are no issues or problems with using central monitor data.

While the CD includes information indicating that people spend 90 % of their time indoors where ozone is substantially attenuated, and examples of indoor/outdoor ratios and examples of personal exposure data (Tables 3-3 and 3-4), the full ramifications of these facts are not discussed in the SP. One ramification that needs to be discussed in relation to biologic plausibility is that the vast majority of the people the vast majority of the time are exposed to substantially below the levels measured at monitors sited to measure peak outdoor concentrations.

For example, a recent paper by Xeu et al.<sup>25</sup> presents information on personal exposures. Data are available from personal exposures of 160 children (1512 6-day measurements) that participated in the Southern California Chronic Ozone Exposure Study. The authors noted that, in the high ozone season in southern California, outdoor and central monitor ozone concentrations were about three times as high as personal and indoor concentrations. In the low ozone season, this ratio was from 5 to 8.

The marked attenuation of personal exposures needs to be considered when evaluating the ozone exposures of the population in relation to the data from clinical or animal studies, as well as in relation to the epidemiologic studies. When evaluating the plausibility of the associations reported in epidemiologic studies, the distribution of personal exposures that accompany any given outside ozone measurement needs to be considered. One ramification that is discussed in the SP is that, based on statistical considerations, these differences should generally bias the results toward the null making it more difficult to identify the true effect. However, as discussed throughout these comments, the full range of individual associations for hospital admissions and mortality per increment of ozone is very large, too large to be biologically plausible. For example, the range of individual-city mortality associations given in Figure 7-29 of the CD is from - 5 to + 10 % per standard ozone increment. Setting aside the issue of negative (or protective) ozone associations, the overall range is much too great to be biologically plausible. If the actual personal exposures associated with these health effects are considered to be 1/3 of those measured at the central monitors, it implies the range is actually -15 to + 30 % per standard ozone increment, which is biologically impossible. Clearly there are many false positives and false negatives in these analyses. This is contrary to the view that all the measurement error issues bias the analyses towards the null. Thus, there are many factors related to the time-series epidemiologic studies that

---

<sup>25</sup> J. Xeu, et al., *J. Air & Waste Manage. Assoc.*, **55**, pages 1508-1515 (2005).

are not yet understood. The SP must fully consider all the ramifications of the attenuation of ozone indoors.

### Confounding by co-pollutants

Section 3.4.2.2 downplays confounding by co-pollutants thereby coming to the conclusion that single-pollutant model results should be the baseline for the quantitative risk assessment. However, there are several major findings from the Stieb et al. 2002, 2003 meta-analyses (cited in the CD) that are particularly relevant to this issue in the SP. Stieb et al. evaluated mortality associations for each of the major pollutants and found a very similar distribution of associations for each in their analysis of 109 separate mortality/time-series studies. For example, the patterns in Figures 1 to 5 of Stieb et al. 2002 are remarkably similar to one another and to the pattern in Figure 7-22 of the CD. They also reported a similar pattern of associations with various categories of death and stronger associations in the warm season than in the cold season for all the pollutants. When multi-pollutant models were compared to single-pollutant models, the associations were reduced in the case of each pollutant. In the case of ozone, the combined association became non-significant in multi-pollutant models. Since various investigators included different pollutants as potential confounders and different modeling strategies, the multi-pollutant analyses available in the literature are not a systematic evaluation of the issue. Nevertheless, the finding that the combined association was reduced substantially for each pollutant when multiple-pollutant model results were compared to single-pollutant results is important.

Taken together, these findings implicate all the pollutants as potential causes and/or confounders of excess mortality in time-series studies, and show that stronger associations in the summer are not restricted to ozone. Since the Stieb et al. meta-analyses like the other meta-analyses are subject to publication bias, the associations reported probably all overstate the true associations. Nevertheless, the pattern of results should not be affected by publication bias the way the magnitude is.

Ozone is the only pollutant for which NMMAPS seasonal results have been published to date, although a seasonal PM10 analysis has been carried out.<sup>26</sup> The NMMAPS database is valuable because, by including all the 90 largest U. S. cities with data, it avoids the issue of publication bias. The wide range of ozone associations in the data, using the same methodology, demonstrates the inherent noise or variability in the data. The same wide pattern of mortality associations ranging from strongly positive to strongly negative is observed in the individual NMMAPS data for all the pollutants studied (PM10, ozone,

---

<sup>26</sup> The study is available as Working Paper #41 on the Johns Hopkins University Dept. of Biostatistics web site. Using updated mortality data from 1987-2000 in 100 cities, the analyses by season show that the combined association at lag 1 was greatest during the summer. Summer was the only season for which the combined effect was statistically significant. An analysis by geographical regions showed a strong seasonal pattern in the Northeast with a peak in the summer and little seasonal variation in the southern regions of the country.

CO, NO<sub>2</sub>, and SO<sub>2</sub>) at all the lags studied.<sup>27</sup> In the NMMAPS morbidity study, only PM10 was evaluated in single pollutant models. However, there was also an implausibly wide range in the 14-city hospital admission associations with PM10 ranging from strongly positive to strongly negative.

The SP and CD dismiss concern over confounding by gases since the peaks of the various pollutants occur in different seasons (page 7-87 of the CD) and because ozone is generally not highly correlated with other pollutants (SP at page 3-34). This is unwarranted since the Stieb et al. meta-analysis shows that the mortality associations for these potential confounders are strongest in the warm period. The cause or causes of these seasonal patterns are not clear. It could be a seasonally varying bias that is yet unidentified. It could be related to toxic constituents (gases or particles) that have a summer maximum. Since it occurs for each pollutant, confounding by as yet not understood weather effects is a likely candidate. Many air pollution analysts dismiss weather effects but the substantial seasonal trends in the health data are caused by subtle changes in weather and human behavior. Therefore, confounding by weather is a candidate. We give an example below of confounding by weather within the summer season that can dramatically affect the results of time-series analyses.

The only reported NMMAPS ozone results in multi-pollutant models were for the summer. In each case, when other pollutants were included, the combined ozone association was reduced and became non-significant. (see Figure 7-28 of the CD) If the combined ozone associations are not robust to inclusion of other pollutants in the summer when there was a small combined positive association in single pollutant models, it is very unlikely that they could be caused by ozone.

This section also overstates the robustness of associations for symptoms by indicating that “in reports from U. S. multi-city studies of respiratory symptoms, associations with ozone were found to remain statistically significant and little changed in magnitude in two pollutant models including PM10 or PM2.5.” (page 3-35, referencing the Mortimer et al. 2002 and Gent et al. 2003 studies at this point in the text) As noted above, the Mortimer et al. study of eight cities is the largest of these studies and the ozone association with symptoms was reduced and became non-significant in every multi-pollutant model tested.

#### Model specification

Although the SP indicates that it is a challenge to distinguish effects due to ozone from other time-varying factors, the SP quotes the CD to indicate that ozone effects are generally robust to various model specifications. However, there is ample evidence that uncertainty due to model selection or model specification is much larger than the SP acknowledges.

---

<sup>27</sup> J. M. Heuss and J. J. Vostal, Comments on the Fourth External Review Draft of “Air Quality Criteria for Particulate Matter” EPA 600/P-99/002aD, June 2003, Prepared for General Motors Corporation. August 28, 2003.

An important systematic analysis was carried out by Koop and Tole 2004<sup>28</sup> who used Bayesian model averaging to evaluate model uncertainty in time series analyses using an extensive set of pollutant and meteorological variables from Toronto, Canada. They summarize their results as follows:

“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.”

Importantly, the authors demonstrate that the results of a single model based on a sequence of hypothesis tests will overestimate the certainty of the results. This is not a new finding in the statistical literature but it has not been carefully considered in the air pollution literature. They use an example to show how the results of a single regression “may lead researchers to make misleading inferences about pollution-mortality effects, thereby seriously underestimating the true uncertainty in the statistical evidence.”

In the GAM re-analysis, the HEI Special Panel also concluded that model selection issues such as specification of weather and degree of control for time “introduce an element of uncertainty that has not been widely appreciated previously.” In fact, the Koop and Tole analysis is the kind of analysis the Panel recommended to investigate the sensitivity of results to model selection issues. By rigorously evaluating the uncertainty with Bayesian model averaging, they show that there is much greater uncertainty in the time-series studies than commonly reported.

While the CD makes note of the Koop and Tole analysis on page 7-16, it is basically dismissed. Because of the issues raised concerning model uncertainty and multiple testing, this paper is much more important than acknowledged in the CD. It raises fundamental questions concerning the utility of time-series studies. These issues need to be acknowledged and discussed in the SP.

Koop and Tole, as noted above, show that individual model results are unreliable. In comments on the federal PM Criteria Document, AIR has presented evidence<sup>29</sup> that led us to same conclusion. By empirically comparing the results of different time series studies of the same city by different investigators, we showed that the results change, often substantively. Subtle differences in model selection can shift the strength of association with a given pollutant, can change the pollutant or pollutants implicated by a given study, and can change the health endpoints that are supposedly affected by the pollutant or pollutants. In the second draft of the PM CD, EPA acknowledged this issue using Philadelphia as an example. There are, however, many more examples in the literature for

---

<sup>28</sup> G. Koop and L. Tole, *J. of Environmental Economics and Management*, **47**, pages 30-54, 2004.

<sup>29</sup> See AIR, Inc. comments on 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> drafts USEPA PM Criteria Document.

ozone and other pollutants as well as for PM. The Agency should evaluate the ozone database to identify cities where there are multiple studies and compare the results.

A recent paper by Keatinge and Donaldson<sup>30</sup> provides important new insights into the issue of modeling weather effects in ozone studies. They evaluated whether mortality that is often attributed to ozone and other pollutants in hot weather results from confounding by neglected weather factors. Their analysis was restricted to days when the mean daily air temperatures exceeded 18 degrees C in Greater London from 1991 to 2002, and evaluated mortality counts at age greater or equal to 65. The adjustment for acclimatization was based on the characteristic pattern that has been reported by various investigators that the rise in mortality on hot days is followed by a prolonged reduction in mortality lasting at least 14 days. When only current temperature (average of days 0 to – 2) was considered in the model, significant mortality was attributed to ozone. When they allowed for cumulative exposure to heat throughout the summer and for sunshine (which contributes to heat stress at any given temperature), the ozone association was reduced by a factor of ten and was no longer statistically significant. For PM10 and SO<sub>2</sub>, summer associations with mortality were reduced and became insignificant when wind speed (which affects both pollutant concentrations and heat stress in the same direction) was added to the model.

This study indicates that previously neglected weather factors may be confounding the mortality analyses relied on in the SP. It is already noted in the CD that variations in treatment of weather can change the results by a factor of 2 and that publication bias can inflate the perceived association by a factor over 3 (page 7-83). The Keatinge and Donaldson analysis suggests that previously overlooked weather factors can reduce the association by a factor of 10. Thus, the fact that the uncertainty due to model selection is much larger than the typical confidence limits on any given statistical association should be highlighted in the SP and considered in the interpretation of the risk assessment.

### Consistency

This section refers to the results of various studies as being fairly consistent or generally consistent and provides only one statistic, that the effect estimates for mortality from several continents range from 0.5 to 5 %. (page 3-37) This summary is highly misleading. The reference to the 0.5 to 5 % range comes from the CD (page 7-139 and page 8-56 ) that describes the range of ozone mortality estimates as being relatively narrow with 80 % of the studies being between 0.5 and 5 % increase per 40 ppb increase in 1-hour maximum ozone. Thus, the SP omits the important qualifier that the range refers to only a portion of the data not the entire range. However, the full range of individual city results is more like - 8 to + 7 % in Figure 7-14 of the CD and - 4 to + 17 % in Figure 7-19. Even this range is misleading, because the Figures present the combined result for NMMAPS (and other multi-city studies) rather than the individual city results which, for summer, varied from – 6 % to + 6 % for a 20 ppb increase in 24-hour ozone. In the Gryparis et al. European multi-city study, the range in yearly analyses

---

<sup>30</sup> W. R. Keatinge and G. C. Donaldson, "Heat acclimatization and sunshine cause false indications of mortality due to ozone, Environmental Research, 2005, in press.

was from  $-12$  to  $+6$  %, and in summer analyses was from  $-8$  to  $+8$  % per standard ozone increment. This wide a range is not biologically plausible. The authors of the strongest association, Vedal et al. 2003, highly qualify their results noting that they are not biologically plausible and that ozone may be acting as a surrogate for something else in the meteorology-air pollution mix.

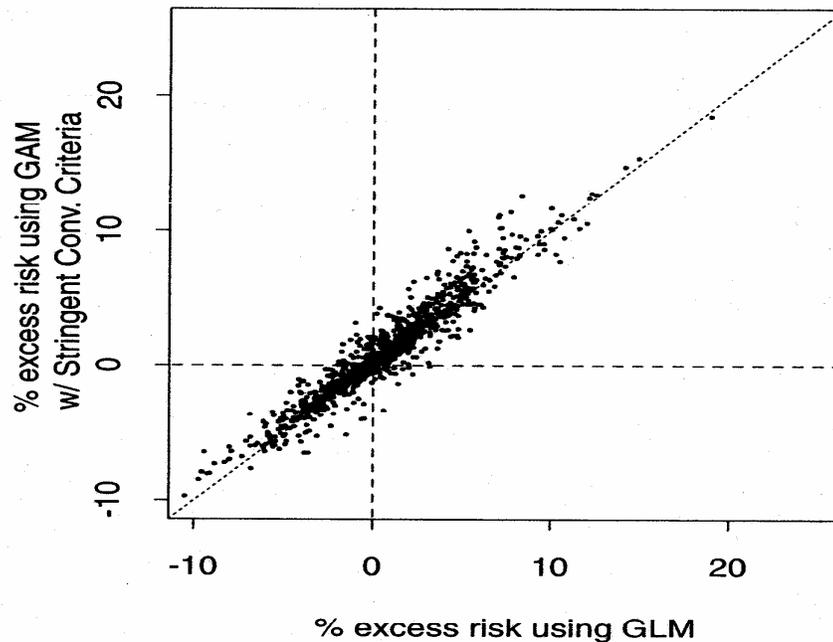
For other health endpoints, as we pointed out above, there is also a very wide range of results ranging from negative to positive. While there are some inverse or negative air pollution associations reported in the literature, the NMMAPS study shows that there are many more negative associations in the data than in the literature. When the GAM issue was raised and many time series studies were re-analyzed, Ito<sup>31</sup> systematically 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 below, in Figure 14 (Figure 2 from Ito 2003) there was a wide range of negative and positive risks in Detroit when all pollutants, lags, and endpoints were considered. Ito also showed (see Figures 3 and 8 in Ito 2003) 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 plot 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.

As we have documented, model uncertainty is much greater than generally appreciated, there is evidence for several major biases in the data in the literature, and systematic studies do not implicate ozone any more than any other pollutant. As EPA reviews the air quality standards for each pollutant in turn, there is some evidence in the literature to implicate that pollutant in premature mortality and other health endpoints. The text spends a great deal of time discussing the warm season ozone associations. However, the Stieb et al. meta-analysis showed that warm season associations are stronger than cold season associations for other pollutants too, so ozone is not unique in that respect.

---

<sup>31</sup> K. Ito, in Health Effects Institute, Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, May 5, 2003.

**% excess risk per 5th-to-95th %ile air pollutants for all outcomes, lags, and air pollutants**



**Figure 14 --Figure 2 from Ito 2003**

Instead of concentrating on the studies in the literature that report positive associations, the authors of the Chapter should address questions such as how there can be such a wide variation of ozone associations both positive and negative in the data. It is not biologically plausible that ozone would be causing mortality in some cities in the summer and protecting against mortality in others. Similarly, it is not plausible that ozone is, on balance, dangerous in the summer and protective in the winter.

Another factor noted in the CD but not in the SP that should be considered is the lack of an ozone mortality signal such as the PM/sulfur-related mortality signal from the 1952 London episode. The lack of a strong health effects signal from the extremely high ozone concentrations measured in Southern California in the 1950's and 1960's argues against a causal ozone mortality association. If a 40 ppb change in 1-hour ozone causes up to a 5 % increase in mortality, as the Chapter suggests, or an even a larger increase as some studies in the literature suggest, the mortality signal in the 1950's and 1960's with peak 1-hour ozone concentrations of 500 to 700 ppb would have been too large to miss. While there was a clear respiratory symptom signal in Southern California on high smog days in the 50's and 60's, there is little evidence of such a signal in more recent years. This is consistent with the clinical studies that suggest that there are thresholds for the first effects, which are neurally-mediated, that, according the USEPA exposure analysis, are

not exceeded for the majority of the population the majority of the time, even at current ozone levels.

### Concentration-response relationships

The SP concludes that because there is insufficient information to support use of potential thresholds, the data should be used within the range of air quality concentrations down to a policy-relevant background in the quantitative risk assessments. (page 3-40) However, the issue of the shape of the concentration-response functions and/or the presence or absence of thresholds in time-series air pollution epidemiology was thoroughly evaluated in the recent PM review and CASAC rejected such an approach for PM. The CASAC PM Panel<sup>32</sup> specifically “did not agree with EPA staff in calculating the burden of associated incidence in their risk assessment using either the predicted background or the lowest measured level in the utilized epidemiological analysis.” The Panel went on to indicate “The available epidemiological database on daily mortality and morbidity does not establish either the presence or absence of threshold concentrations for adverse health effects.” In order to avoid overemphasizing an approach that assumes effects that extend to the predicted background concentrations, the CASAC panel recommended use of a specific assumed threshold as the primary analysis.

In the case of ozone, all the same methodological issues apply. In particular, the issue that measurement error can mask the presence of a threshold is very important. In addition, however, there is much more skepticism over whether the associations are causal or not. Therefore, a range of assumptions concerning causality and the shape of the concentration-response function should be evaluated in the risk assessment.

### **Section 3.5 Biological plausibility and coherence of evidence**

This section draws upon Chapter 8 of the CD. However, that chapter will be revised based on CASAC and public comments and CASAC felt the revisions were of such import that Panel reserved the option of holding a teleconference to discuss providing additional advice to the Administrator after the final CD is published at the end of February 2006.

In discussing the plausibility and coherence for each category of potential effects, this section overstates the case. For respiratory morbidity, there is no question from controlled human studies that exposures to elevated ozone along with moderate to heavy exercise causes transient and reversible respiratory effects. However, the doses/activity levels that can cause these effects do not routinely occur in the population at the level of the current ozone standard. Therefore, it is not clear whether the epidemiological findings for hospital admissions or emergency department visits are causal. The widely inconsistent and therefore inconclusive results for emergency visits raises a serious

---

<sup>32</sup> EPA’s Review of the National Ambient Air Quality Standards for Particulate Matter (Second Draft PM Staff Paper, January 2005), A Review by the Particulate Matter Review Panel of the EPA Clean Air Scientific Advisory Committee, EPA-SAB-CASAC-05-007, June 2005, Letter from Dr. Rogene Henderson to Administrator Johnson, page 6.

question as to how to interpret the respiratory hospital admissions studies. The fact that there is no consistent pattern in specific categories of admissions together with the overall wide range of positive and negative associations in systematic analyses argues against relying on these studies to portray risk.

For effects on the cardiovascular system, the lack of any substantive effects in controlled human studies together with the low personal exposures of the population on a day-to-day basis argue against any cardiovascular effects.

For long-term effects, the low long-term personal exposures of the population together with the lack of consistent effects in the over 40 chronic studies reviewed in the CD argue against chronic ozone effects.

For mortality, the lack of any mechanisms by which the transient and reversible irritant effects of ozone could cause mortality together with the low personal exposures of the population on a day-to-day basis argue against any causal mortality effects. The SP argues that the pattern of effects suggests causality, but the overall wide range from positive to negative in multi-city or systematic studies argues against causality. In addition, the same overall pattern of results is seen for each pollutant and publication bias leads to a false appearance of strength and consistency in the epidemiologic database.

All the above facts and arguments need to be added to this section to provide a balanced discussion of the issues of biologic plausibility and coherence.

### **Section 3.6 Ozone-related impacts on public health**

The key issue in this section is how to evaluate the adversity of the acute respiratory responses to ozone. This was a major point of discussion at CASAC during the previous review and the staff includes tables describing the results of that discussion in this review (Tables 3-2 and 3-3).

During the previous review, EPA staff and CASAC spent a considerable amount of time evaluating the current understanding and divergence of opinion in the scientific community as to what respiratory effects and degrees of response might be regarded as adverse health effects associated with ozone.<sup>33</sup> They considered asthmatic and healthy individuals and categorized the various functional and symptomatic responses. Based on discussions with medical experts who have worked with asthmatics, staff concluded that single ozone exposures that resulted in moderate responses are not likely to interfere with normal activity nor to result in increased frequency of medication or the use of additional medications. However, staff felt that moderate exposures when repeated could result in an increased likelihood of many asthmatic individuals to limit normal activity. Therefore, they recommended that moderate symptom and functional responses, when repeated, should be considered as adverse health effects. For healthy individuals, there was a consistency of CASAC opinion that single, acute moderate health responses should not be considered adverse. Because of concern over repeated health effects, staff

---

<sup>33</sup> See 1996 USEPA Staff Paper at pages 59-72.

recommended that the number of ozone exposures resulting in moderate health effects should be considered as a factor in characterizing the adversity.

Since that review, there is now more information on the mechanism of the first respiratory effects of ozone. In addition, there are now several studies that found no long-term lung function changes that could be ascribed to ozone in children in the high ozone areas of California.

Ozone is well established as an irritant gas. The CD includes discussion of several of the body's defense mechanisms against ozone. It points out that antioxidants are first line of defense against oxidative stress. It acknowledges that a neurally-mediated mechanism is responsible for the lung function decrements and symptoms seen in clinical studies. Although it is well established that ozone causes transient, reversible lung function decrements at sufficient ozone doses, the evidence argues against the presence of chronic lung function decrements. Chapter 6 of the CD describes the body of studies that show that the first effects of ozone are caused by reflex responses of vagal nerve receptors that involuntarily inhibit maximal inspiratory effort. This response is the body's reflexive reaction to the presence of an irritant gas and is unrelated to sensations of discomfort. The CD acknowledges that most effects are transient and of a reflex nature and are not particularly life threatening (page 8-70).

Given more knowledge of the reflex cause of the lung function effects and the lack of findings on long-term lung function changes associated with ozone in the Southern California Children's Health Study, there should be less of a concern for repeated acute effects over the ozone season that might lead to irreversible long-term changes. EPA staff and CASAC should consider these factors in determining which physiological effects should be considered as adverse as well as the public health significance of those physiological effects.

### **Section 3.7 Summary and conclusions for ozone health effects**

This section significantly overstates the consistency, plausibility and coherence of the epidemiological evidence. Although there is now a distinctly larger data base of epidemiological studies related to ozone, a number of critical findings in the epidemiological literature - some of which are noted in the CD and some of which are not - are omitted or downplayed in this section. These include the fact that model selection is now acknowledged as a much more important issue that it was in 1996. The CD notes that it can change associations by a factor of two; we included an example where using previously ignored weather factors reduced a positive association with ozone and mortality in the summer in London by a factor of ten.

Publication bias is receiving greater attention now and the comparisons noted by Goodman and included in the CD indicated that it reduced the ozone/mortality association by a factor of more than three.

The CD also notes that findings of associations with respiratory emergency room visits now are viewed as inconclusive since there are many studies that report negative or insignificant associations with ozone.

Among the important facts that are omitted in the CD and SP are the presence of an extremely wide range in associations from strongly negative to strongly positive in multi-city studies of mortality, and the presence of similar patterns of association for all the major pollutants with mortality and morbidity in systematic studies such as NMMAPS, Stieb et al. 2002, 2003, and Ito 2003. In addition, the presence of stronger associations in the warm season is not unique to ozone; it has been reported for all the major pollutants.

One finding that is prevalent for ozone, is the presence of generally negative associations in the winter. The presence of negative associations for ozone is not limited to the winter, however. In both NMMAPS and the Gryparis et al. European multi-city study, there is a wide range of negative to positive ozone associations for the summer also.

When these facts that are downplayed or omitted in the SP are taken into account, the uncertainty in the epidemiologic studies is at least as great as it was in 1996. The current draft CD already uses vague language to describe the database for many of the endpoints investigated using words like “suggests” or “limited evidence” or “inconsistent.” When additional uncertainty due to model selection, publication bias, and potential confounding from other environmental variables (weather and pollutants) is considered, the conclusion that the acute time-series associations are likely causal cannot be supported. The small lung function decrements reported in field studies of active children are likely causal but the presence of a consistent association with symptoms or medication use is less clear. For chronic effects, the CD already acknowledges that “the strength of the evidence does not allow establishment of a likely causal relationship between chronic ozone exposure and these health outcomes” (page 8-58).

## **Chapter 4 Characterization of Human Exposure to Ozone**

This chapter introduces and presents the results from the Air Pollutants Exposure (APEX) model that staff is using to estimate ozone exposures and activity levels for 12 urban areas. It is used in Chapter 5 to link to dose-response functions from the controlled studies and estimate the risk of acute respiratory effects. The controlled studies provide a body of information which has been used in the past to set ozone air quality standards and still provide the best available data on ozone effects. However, the information on effects in controlled studies needs to be considered in relation to the distribution of human exposures that occurs in the U. S. In the 1996/97 review, that translation involved application of the pNEM model along with dose-response relationships drawn from the controlled human studies. The model was used because subjects need to be outside and exercising at the time and at the place of high ozone to be at risk from the respiratory effects found in controlled studies. The model was used to calculate the likelihood of risk from single and repeated exposures.

Over the past decade the Agency has refined the ozone exposure modeling as detailed in the Chapter, the draft Technical Support Document that was made available in November 2005, and in various reports and publications referenced in these documents. The 1996 exposure and risk analysis, as documented in the 1996 Staff Paper, had many conservative assumptions in both the exposure/activity levels calculated and the way the dose-response functions were implemented in the model. For example, the analysis used 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<sup>34</sup> that the analysis allowed more high ventilation rates than would actually occur in the populations of interest - outdoor workers, outdoor children, etc. It is important that all the assumptions and inputs in the new exposure and risk analysis be carefully evaluated to insure that it does not materially overstate the risk.

In addition, the results should be shown for the entire distribution of exposures not just the extremes so that the plausibility of the acute epidemiologic associations being real health effects can be evaluated. The SP glosses over several important facts concerning ozone exposure that are critical to interpreting the epidemiological studies and evaluating them in light of the clinical and animal studies. The CD (Table AX3-2) provides pooled distributions of the ozone levels from monitors across the country. One statistic relevant to the interpretation of time-series studies is the mean or median concentration. The median ozone concentration in Metropolitan Statistical Areas (MSAs) is 0.032 ppm. Table AX3-2 also indicates that the mean concentration is 0.033 ppm and that 90 % of the daily average ozone levels monitored in urban areas across the country are below 0.050 ppm. These data are from the May through September ozone monitoring season. Thus, even though many of these areas are out of attainment of the ozone standard, the typical day-to-day ambient ozone levels in cities are currently very low. In fact, they are typically within the range of levels of policy relevant background shown in Table 3-2 of the CD. As shown in Figure 4 above, the ozone season mean concentrations are very similar all across the country in both heavily-populated urban area and sparsely-populated rural and remote areas.

While the SP in this chapter emphasizes the factors affecting ozone exposures, it does not document what the full distribution of personal exposures is especially relative to the monitoring data. The reader is not informed that personal exposure monitoring confirms that the exposure to ozone in the breathing zone of individuals, which is the proximate cause of any ozone health effects, is typically only a fraction of the ozone level measured at monitors. The CD also documents that indoor exposures are typically only a fraction of outdoor exposures. Importantly, the reader is not informed that the ozone monitors, in many instances, are located outside the breathing zone of the general population. These are important facts that are well documented in the CD and are crucial to the evaluation of whether current ambient exposures can be causing premature mortality and to the proper estimation of the full distribution of personal exposures in the population. Both

---

<sup>34</sup> See U. S. Environmental Protection Agency, Review of National Ambient Air Quality Standard for Ozone: Assessment of Scientific and Technical Information, OAQPS Staff Paper, EPA-452/R-96-007, June 1996 at page 102.

Chapter 3 and Chapter 4 of the SP need to be revised to present and discuss the implications of these facts.

With many examples of the personal exposure to ozone being less than half of that measured at central monitors, the average exposure of the population of 0.033 ppm translates into an average personal exposure of 0.016 or less during the ozone season. While personal exposures may be higher for some portions of the population, such as outdoor workers, personal exposures of the frail population, the elderly, will tend to be less than this. These findings concerning personal exposure need to be included in the Chapter and they need to be explicitly considered throughout the SP. The fundamental question is whether day-to-day changes in ambient ozone within the range of 0.01 to 0.06 that are associated with personal exposures that are within the range of 0.005 to 0.04 ppm are capable of causing or contributing to premature mortality.

During the December 2005 CASAC meeting, the question of how the performance of the APEX model can be or should be evaluated and validated was raised several times by CASAC panelists. This was also a question raised relative to the pNEM model during the previous review. Johnson et al. 1996<sup>35</sup> conducted a pilot study of personal exposures in Southern California and reported that the mean ratios of monitored to modeled personal exposures were between 0.5 and 0.6, indicating that pNEM systematically overestimated actual personal exposures. Since many of the parts of the pNEM model are being used in APEX, it is incumbent on the Agency to demonstrate that the APEX model does not materially overestimate personal ozone exposures either in the aggregate or in the upper tail of the distribution.

Thus, it will be particularly important for the Agency to evaluate the APEX model's performance versus real personal exposure data. In addition, the Agency should seek to evaluate the model's intermediate outputs for key variables that influence the final result. These include the distribution of time outside as well as the distribution of equivalent ventilation rates for the subjects simulated in the model.

The upper tail of the distribution of equivalent ventilation rates is particularly important because it is a critical factor in determining the frequency of exposures of concern. The exercise utilized in the 6-8 hour controlled exposure studies was intended to simulate work performed during a day of heavy to severe manual labor by outdoor laborers (McDonnell et al. 1991). Although the ventilation rate (40 L/min) is described as "moderate exercise," activity maintained at this level for 6 to 8 hours has to be considered "heavy" or "strenuous work or play." Therefore, it is particularly important to evaluate how the model's predictions of the number of occurrences of elevated ventilation rates compare with real world data.

Another important factor that is documented in the CD but not included in the APEX model is that ozone is often measured at heights between 3 and 15 meters and people's

---

<sup>35</sup> T. Johnson, K. Clark, K. Anderson, A. Geyh, and W. Ollison, "A Pilot Study of Los Angeles Personal Ozone Exposures During Scripted Behaviors," in *Measurement of Toxic and Related Air Pollutants*, VIP-64, Air & Waste Management Association, Pittsburgh, PA, pages 358-365, 1996.

exposure outdoors occurs at 1-2 meters. (see page 3-15 and page AX3-190 of the CD) While the reduction in ozone near ground level is discussed in terms of plant exposures in the CD, it is also particularly relevant to human exposures. For example, Wisbeth et al.<sup>36</sup> measured the increment between ozone at 2 and 10 meters and reported an average 13 % 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-other microenvironment. The SP indicates that an analysis of the uncertainty due to interpolation errors will be provided in the next SP. (page 4-24) However, this just involves the horizontal variability. The uncertainty due to variability both horizontally and vertically should be evaluated and included in the next SP.

APEX calculates equivalent ventilation rates (EVR) and then puts the data into bins of moderate (12-27 EVR) and heavy exertion (> 27 EVR) as noted in Table 4-1. (The bins are from Whitfield et al. 1996). Whitfield constructed dose-response functions from EPA controlled exposure studies that used a single nominal breathing rate, 40 L/min. However, the subjects were of different sizes and their equivalent ventilation rates (normalized to body surface area) averaged 20 L/min/m<sup>2</sup> and ranged from 13-27. When APEX calculates the EVRs, the current model lumps all EVR exposures between 13 and 27 into one bin and assumes that the FEV decrements are those measured at 20 EVR. (The details of the APEX dose-response functions are not included in the material made available for public review. However, the SP indicates that a similar methodology was used to that developed for the previous risk assessment, page 5-13 of the SP)

The 1996 CD showed that 1- to 2- hour ozone responses vary with EVR and Ollison has shown that the 8-hour responses in the Folinsbee/Horstman data do too.<sup>37</sup> Thus, the dose-response curve varies with EVR. The distribution of exposures also varies with EVR. We expect many more 8-hour occurrences of ozone exposures at 13 EVR than at 20 and many more at 20 than at 27. Thus, the approximation that all exposures between 13 and 27 EVR respond as though they were at 20 EVR will substantially overestimate the risk.

Since the model calculates EVR to a much finer scale, the distribution of EVR should be presented in the SP and dose-response curves should be developed as a function of EVR to a void the over-estimation of risk.

---

<sup>36</sup> 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 89<sup>th</sup> Annual Meeting of the Air and Waste Management Association, Nashville, TN, June 1996.

<sup>37</sup> Appendix A to American Petroleum Institute comments on EPA's December 13, 1996 proposed ozone rule.

Even with likely over-estimates of the number of exposures, the APEX results presented in Tables 4-7 to 4-12 demonstrate that the current standard effectively limits the number of persons and person-occurrences of ozone exposures to 0.08 ppm for 8-hours with moderate exertion (strenuous work or play). Only a fraction of a percent of the different groups experience an exposure of concern during the year (Table 4-9) and only about one percent of the fraction of a percent experience more than one exposure of concern in a year (page 4-26). Even for historical 2004 air quality, the average number of exposures above 0.08 ppm for individuals is 1.2, which is much less than two per year.

## **Chapter 5 Characterization of Health Risks**

Chapter 5 presents the initial results from an ozone health risk assessment that that considers lung function decrements, hospital admissions, and mortality. The results of controlled studies along with the detailed personal exposure estimates from the APEX model are used to characterize the risk of lung function decrements. Selected statistical associations from the epidemiologic database are used together with ambient monitoring to estimate the potential hospital admissions and mortality effects. Since these two approaches are vastly different, we provide comments separately on each calculation. For lung-function decrements, staff should re-visit the shape of the dose-response curve allowing for non-linear responses and including new data from low concentration experiments (below 0.08 ppm). The use of linear dose-response is problematic since evaluations of the FEV and other acute respiratory responses have been found to be non-linear. The use of a linear response will overestimate the risk based, not on data of effects, but based on an unsupportable assumption. In addition, the full details of the methodology need to be made available to both CASAC and the public.

For mortality and hospital admissions, the risk assessment should include explicit consideration of a range of assumptions concerning causality, background levels, and shape of the dose-response. In addition, the full range of associations in systematic analyses from positive to negative should be included to evaluate the uncertainty in the epidemiologic data base and the implications of both protective and harmful effects of ozone that vary with study, city and season.

### **Lung function decrements**

From the extensive data on ozone respiratory effects in controlled studies, decrements in the performance of lung function tests are clearly caused by ozone. In addition, we have a great deal of knowledge concerning the mechanisms behind these effects and the influence of key exposure variables on these effects. We also have the output of the APEX model to estimate exposures. However, there is still uncertainty in the risk estimates that come from (1) the APEX model itself, and (2) the way the dose-response functions from the controlled studies are developed and applied to the APEX output. As our comments on Chapter 4 demonstrated, APEX has not been validated or even fully evaluated. Since its predecessor, pNEM, overestimated the risk of respiratory effects in the last review (as acknowledged in the 1996 SP and shown by Johnson et al. 1996 and American Petroleum Institute 1997), we are concerned that APEX overestimates the

number of occurrences of elevated 8-hour ozone exposures with strenuous work or play. It has not been possible, in the short time available for public review, to fully evaluate all the inputs and assumptions in the APEX model.

Although the SP indicates (page 5-13) that a similar methodology to that developed in the prior risk assessment was used to develop the dose-response relationships for 8-hour ozone exposures, all the details of that methodology are not publicly available. One of the key assumptions in the analysis is acknowledged to be the method to extrapolate the responses below the lowest exposure levels used in the controlled studies. The SP indicates that a linear extrapolation down to background levels was assumed. (page 5-15) During the December 2005 CASAC meeting, EPA staff showed a plot of the dose-response function for 10, 15, and 20 % decrements. The plot showed that 20 % decrements were predicted at 0.05 ppm and above and that 10 % decrements were predicted down to about 0.02 ppm. The SP also indicates that the responses were corrected for the effect of exercise in clean air, typically an average improvement of several percent. It is not clear how this is implemented in the model. If the baseline is several percent improvement, then the dose-response shown to CASAC would predict 20 % decrements below 0.05 ppm with exercise. With clean air exposure there is also random variation and measurement error, so that there are both decrements and improvements of up to 10 % in the clean air exposures.<sup>38</sup> It is not clear how these factors are considered in the EPA analysis. The full details of the methodology need to be made available to both CASAC and the public.

The use of linear dose-response is problematic since evaluations of the FEV and other acute respiratory responses have been found to be non-linear. Thus, the use of a linear response will overestimate the risk based, not on data of effects, but based on an unsupportable assumption.

McDonnell et al. 1997 point out that the lung function response is non-linear in each of the three exposure variables -- concentration, ventilation, and time. They developed a sigmoid-shaped model for ozone induced- FEV changes from 2-hour exposure data. In addition, Mudway and Kelly 2004 modeled the inflammatory response in acute ozone human chamber studies and found a threshold for the effects when they used a linear CVT model. As noted during the December CASAC meeting, the fit of the inflammation data with the CVT model could be improved by use of non-linear models.

The use of a linear dose-response for lung-function decrements was questioned by several CASAC panelists in December. It was also indicated during the meeting that a paper is in press that reports the results of ozone exposures below 0.08. Therefore, staff should revisit the issue of the shape of the dose-response including the new data and revise the methodology for the second draft SP. In addition, the detailed methodology must be made available to CASAC and the public to allow for a proper review.

---

<sup>38</sup> W. F. McDonnell, et al., Am. J. Respir. Crit. Care Med., **156**, 715-722, 1997.

The detailed review of ozone respiratory effects and mechanisms by Mudway and Kelly 2000 points out that ozone has a background generally between 20 and 40 ppb. They also note that because it is a highly irritating gas, there is only a small difference between levels viewed as toxic and that of background. They indicate that there is only a factor of 3 or 4 between levels that are detrimental and the natural background.

Mudway and Kelly indicate that ozone is a highly reactive gas that is consumed by reactive processes on reaching the first interface in the lung, the lung lining fluid compartment. They note that reactions between ozone and antioxidants tend to dominate in this compartment and these are generally thought as beneficial or protective interactions. In their review, they discuss a range of antioxidant defenses. They further conclude that when ozone is relatively low - probably less than 80 ppb - the majority of ozone will be neutralized by reactions with uric acid in the nasal passages and upper airways.

As the dose is increased, the reflex response noted above and then immune system responses begin to come into play. Even these immune system responses which are noted as the first indications of “inflammation” are physiological processes that occur in all living organisms under the stimuli of daily life. The first reported changes (that occur with heavy exercise after 1- to 3-hours above a threshold of 0.18 to 0.20 ppm) are well within the range of physiological variability and should not, therefore, be interpreted as unequivocal evidence of ozone adversity. They fall into the category of biochemical markers that ATS indicates do not necessarily imply adversity. For example, Mudway and Kelly note that for neutrophils transiting into the lung – one of the earliest of these responses – it is not clear if the response should be considered beneficial (functioning to clear necrotic cells) or detrimental (leading to an active inflammation with tissue injury).

Based on our understanding of the way the body defends itself against ozone, an irritating gas, and the non-linear responses in human chamber studies, the use of a linear response down to background cannot be defended.

### **Mortality and hospital admissions**

For these endpoints, there are four important issues or sources of uncertainty noted in the SP that we wish to provide comments on. In contrast to the acute respiratory effects where we know the effects are real and the issues are extrapolation below the range of the data and the clinical and public health significance of the effects, the first and main issue in interpreting the epidemiologic associations is whether they are real health effects or not. Although there clearly are positive associations in the literature for these health endpoints, the overall pattern of associations is not biologically plausible. In multi-city studies, there is an implausibly wide range of associations from strongly positive in some cities to no effect in some cities to strongly negative (implying protective effects) in some cities. The full range for several multi-city studies is shown below in Table 1.

**Table 1 Range of ozone individual-city single-pollutant mortality associations in multi-city studies per standard ozone increment as defined in the CD (subject to model selection uncertainty but not publication bias)**

NMMAAPSII (warm season)	- 6 % to + 6 %
Gryparis, et al. (warm season)	- 8 % to + 8 %
Gryparis et al. (year-round)	- 12 % to + 6 %
Bell et al. 2004 (warm season)	not reported, but large (see CD Figure 7-29)
Huang et al. 2005 (warm season)	- 3 % to + 8 %

With this range of heterogeneity, the practice of using Bayesian techniques to shrink the city-specific estimates towards the overall mean is highly questionable. The individual-city estimates are dramatically pulled towards the mean and this masks the heterogeneity, especially when the heterogeneity includes both no effect and protective effects.

As noted above, the SP fails to acknowledge that the body of studies since 1996 demonstrates major uncertainties and issues related to model selection and publication bias. When the full range of studies in the CD is viewed, the range is also very wide but shifted to more positive findings. This is shown below in Table 2. The heterogeneity implied by these ranges is biologically implausible. While the input from the panelists in the December 2005 CASAC consultation varied regarding the use of these data, there was much more skepticism regarding causality than was the case in the PM review. The risk assessment, therefore, should include a probabilistic consideration that the associations may not be causal, particularly for cardiovascular endpoints.

**Table 2 Range of ozone individual-city single-pollutant associations in studies included in CD Figures per standard ozone increment as defined in the CD (subject to model selection uncertainty and publication bias)**

Figure 7-8 Emergency visits for asthma	- 20 % to + 90 %
Figure 7-9 Respiratory hospitalizations	- 17 % to + 17 %
Figure 7-13 Cardiovascular hospitalizations	- 12 % to + 35 %
Figure 7-14 All-cause mortality year round	- 8 % to + 7 %
Figure 7-19 All-cause mortality by season	- 4 % to + 17 %

A second major uncertainty listed by staff is that related to adequacy of ambient monitors as surrogates for population exposure. (page 5-28) The staff indicates that the fact that ambient concentrations overestimate personal exposures results in effect estimates that understate the underlying concentration-response relationships. The text goes on to indicate that the risk estimates in the SP may therefore underestimate the overall impact of ozone on mortality and hospital admissions. This is not true, because the concentration-response functions are being applied to the ambient concentrations that were used in their development.

But it is true that the underlying personal exposure-response relationships implied by these associations are stronger by a factor 2 or 3 than the concentration response relationships suggest. While this does not change the risk calculated in the risk assessment, it does indicate that the ranges of results in multi-city studies is actually 2 to 3 times larger (both positive and negative) than that indicated in Table 1. Such a wide range is biologically impossible.

In addition, the fact that the underlying personal exposure-response relationships implied by these associations are stronger by a factor 2 or 3 than the ambient concentration-response relationships provides an even greater challenge to explain how low ozone concentrations, within the range of background, can be causing the premature mortality estimated in the risk assessment. If one takes the ozone associations at face value as evidence of effects, then one also has to explain how ozone can be both damaging in some cities and protective in others as well as damaging, on balance, in the summer and protective, on balance, in the winter. This pattern is not plausible and is not adequately acknowledged or discussed in the SP.

A third major area of uncertainty noted by staff is a set of issues related to the empirically-estimated concentration-response relationships. These include uncertainty related to model selection, shape of the response, and confounding. As we document above, the uncertainty in the "true" relation is much larger than the statistical uncertainty in any given association. While staff concludes that a linear response down to background should be applied, there are many reasons why this should not be the baseline estimate. As noted earlier in these comments, CASAC rejected this approach for PM even though the PM database was viewed as more consistent. From the controlled human studies, we know that the body has mechanisms to deal with oxidative stress and the first responses are non-linear, transient, and reversible. Without strenuous work or play for several hours or more, these responses are not found. Respiratory symptoms in healthy children were not included as an endpoint in this review because the CD concluded that there was no consistent evidence of an association between ozone and respiratory symptoms in children. (page 5-5) The CD also concluded that the evidence for respiratory emergency room visit is inconsistent and inconclusive as is the evidence for cardiovascular hospital admissions. Assuming that there are mortality effects down to background violates the coherence guideline from toxicology whereby lesser effects are seen at lesser doses and greater effects at greater doses. For all these reasons, mortality effects and any effects down to background should be severely discounted.

The fourth major uncertainty relates to the use of the GEOS-CHEM model results for policy-relevant background. As documented above in comments on Chapter 2, the model will systematically underestimate the peak background levels. Therefore, several alternative higher background assumptions should be used. They could include a constant 40 ppb background, scaling the seasonal pattern from the Trinidad Head data, and a probabilistic formulation that provides yearly peak 8-hour values in the range of 60 to 75 ppb.

As a result of all these uncertainties, the risk results presented in Tables 5-5 to 5-12 severely miss-characterize the certainty with which the risk is known. Since there are now more issues and inconsistencies related to the interpretation of the time-series studies than there were in 1996, when the decision was made not to rely on ozone/mortality associations, a similar decision is warranted for this review. The reader should be made aware of the full range of mortality associations in the literature, and that this range is not biologically plausible. Comparisons of multi-city results with meta-analyses have demonstrated a major issue with publication bias as noted by Goodman. Goodman's caution about publication bias and the current weaknesses of single-site reports along with Koop and Tole's caution about the use of point estimates from time-series to set regulatory standards need to be carefully considered by CASAC and EPA.

## **Chapter 6 Staff Conclusions and Recommendations on Primary Ozone NAAQS**

After the prior chapters are revised in response to CASAC and public comments, the case for health effects below the current standard will be dramatically altered. As we have shown, the acute respiratory effects below the current standard are overstated primarily because of the assumption of a linear response and the assumption that all predicted EVR occurrences above 13 respond as though they were at 20 EVR. A careful evaluation of the APEX model may identify other systematic biases in the results.

While the search for the study that reports an effect at the lowest level will identify real effects in controlled exposure studies (if they can be replicated), in the time-series epidemiological literature, the search will identify outliers rather than real effects. For example, the Vancouver study by Vedal et al. 2003 is such a study but the authors discount the interpretation of the association they report as a causal ozone effect.

Because of the wide range in individual-city results for mortality in NMMAPS, the reader should be informed that the range of effects from "as is" ozone is not 0.1 to 1.9 % as noted on page 6-13 or - 0.9 % to + 3.1 % when the confidence intervals in Table 5-7 are considered, but actually more like - 5 % to + 5 % when the individual-city data from NMMAPS and other multi-city studies are considered. Because of the many issues related to interpreting these associations as causal, the reader should be informed that epidemiology suggests effects that are small and inconsistent and not likely causal for mortality. For respiratory morbidity there are small effects on FEV for active subjects and small and inconsistent effects on emergency room visits and hospital admissions. In

particular, there is a lack of findings for respiratory admissions in southern California that is surprising given the high ozone levels in that area.

The reader should also be informed that, for all the endpoints investigated in time-series studies, there are many similar associations with other pollutants besides ozone. In fact, where systematic analyses have been carried out, as in NMMMAPS, Stieb et al. 2002, 2003 and Ito 2003, similar patterns of associations are reported for many pollutants. This includes a warm season effect.

In contrast to the focus on quantifying the potential mortality effects of ozone exposure in the CD and SP, the CD dismisses consideration of the possible beneficial effects of tropospheric ozone on UV-induced health effects concluding that the beneficial effects cannot be assessed quantitatively at this time. However, the health effects involved are substantial health issues and the causal link to UV is much better established than is the case for ozone mortality in the CD and SP. Section 10.2 of the CD stresses the uncertainties and the limitations in estimating beneficial effects generally characterizing the impacts as small and uncertain. While this may be true, it is also true of the adverse mortality effects implied by many of the epidemiological associations in the CD. EPA's response<sup>39</sup> to the court's remand to consider UV-related effects of ground level ozone discussed various estimates of these effects (made by EPA staff, EPA consultants, and staff at other government agencies) at some length. In all cases, there is no disagreement with the fact that there will be an effect, but the Agency raised questions concerning the estimates in terms of the magnitude of the effect or the certainty with which it can be estimated. We have no quarrel with the decision not to estimate beneficial effects at this time, but the Agency should discuss the evidence for both adverse and beneficial effects with the same objectivity.

Section 6.3.5 indicates that, after viewing the entire body of evidence, it is staff's view that it is appropriate to conduct additional exposure and risk assessments down to an alternative standard level as low as 0.06 ppm. However, in the 1996/1997 review, an 0.07 ppm standard was considered too close to background by both the Administrator and CASAC. What has changed? First, staff stresses one global model application that suggests a low background. But one uncertain model that admittedly doesn't include all the pertinent physical phenomena that influence the extremes isn't sufficient reason. Second, staff points to acute epidemiologic studies, but as we have shown in these comments, staff exaggerates their consistency, ignores the strong evidence of publication bias, and stretches to explain biologic plausibility for severe health outcomes. While analyses can be carried out at 0.06 ppm or lower, as shown in Figure 1, an 0.06 ppm standard with the current statistical form would put virtually the entire country out of attainment including areas that are far removed from sources of man-made precursors.

## **Chapter 7 Policy-Relevant Assessment of Welfare Effects Evidence**

The introduction to Chapter 7 indicates that, of the various categories of welfare effects, the effects of ozone on vegetation are of most concern at concentrations typically

---

<sup>39</sup> 68 Federal Register 614, January 6, 2003.

occurring in the U. S. Therefore, the chapter properly focuses on the ozone effects on vegetation. Because of time and resource limitations, this draft is not as far along in its update of the welfare-related exposure, risk and benefits assessments conducted in the last review as it is for the health-related assessments discussed above. Thus, the draft discusses the plans for the assessments rather than the results of the assessments.

## **Section 7.2 Effects on vegetation**

The draft SP indicates that the science published since the last review in 1996 has not fundamentally altered the understanding and conclusions concerning ozone effects on vegetation and notes that only a few new studies focus on addressing the data gaps or uncertainties identified in the last review (page 7-2). There are, however, a number of findings and conclusions from the data, including the more recent data, relevant to the policy decisions regarding secondary ozone standards.

First, the SP indicates that the open-top chamber (OTC) technique, first developed in the 1970s and applied extensively in the 1980s in the National Crop Loss Assessment Network (NCLAN) is currently the most useful data in a policy context. (page 7-6) However, as noted in the SP, there are several characteristics of the OTC design and operation that can lead to unrealistic exposures. (page 7-4) The SP notes that there are differences in turbulence, microclimate, vertical concentration gradients and water and nutrient availability between the OTC exposures and natural exposures. In addition, to develop concentration-response (C-R) functions, the OTC experiments utilize a charcoal-filtered clean air control, ambient ozone, and ozone augmented in the chambers at levels above ambient. The differences between OTC exposures and natural exposures can lead to systematic biases and/or uncertainty in applying the C-R functions to ambient data.

Second, based on an understanding of the physiology and cellular processes within plants, that has been further clarified and enhanced in the past decade, plants have several defense and repair mechanisms. There are mechanisms that limit entry of ozone into the plant (stomata), a range of anti-oxidants that detoxify ozone, release of ethylene that reacts with ozone, and other defense and compensation mechanisms. (discussed at pages 7-6 to 7-12) The presence of these defense and repair mechanisms indicates that ozone would be expected to display threshold or non-linear responses to ozone exposure.

Third, ozone responses vary with genetic, biological, chemical, and physical factors. Because of genetic differences, species vary greatly in their responsiveness to ozone. The SP notes that even within a given species, individual genotypes or populations can vary significantly in ozone sensitivity. (page 7-13) Temperature, moisture, and nutrients all affect ozone sensitivity. A number of these factors also influence ozone uptake. The combined effect of these factors adds uncertainty when the OTC results with specific cultivars are translated into effects with current cultivars in natural conditions that differ from the conditions present in the original experiments. The SP explicitly acknowledges this uncertainty by indicating that “current understanding of the important components of exposure in eliciting plant response are based on exposure regimes that favored ozone uptake, contained large numbers of peak concentrations, and closely controlled other environmental factors. In the absence of further study, it is unclear how well indices

selected on this basis perform under different exposure and growth scenarios.” (page 7-22)

In addition, the SP notes that the presence of ozone is only one of a multitude of factors influencing the likelihood of an adverse vegetation effect from the presence of the pollutant. (page 7-17) The SP also indicates that it is not possible to predict the impact of modifying factors on the rate of uptake and plant response to that uptake. The SP concludes that “thus any ambient air quality exposure index will of necessity be a simplification of the actual relationship between pollutant concentrations in the ambient air and plant response.” (page 7-17)

Fourth, the SP acknowledges (page 7-17) that most of what we know comes from controlled experiments that sought to minimize the impact of confounding variables so that a clear ozone signal could be measured. It indicates that experimental ozone exposures were typified by the episodic occurrence of a large number of higher ozone concentrations and that applying these results may over- or under-estimate effects in locations where a different temporal pattern is prevalent. Based on the experiments, exposure indices that accumulate exposures and weight the higher concentrations had better statistical fits than either mean or peak indices. However, the SP acknowledges that no experiments were conducted to test the performance of these indices in the field. (page 7-18)

Fifth, the SP acknowledges that there are still important issues related to choosing the most appropriate concentration-based forms. These include the importance of peak versus mid-range concentrations and the importance of night versus daytime exposure.

Sixth, the SP notes the focus in recent years on flux-based models and indices to better relate ambient ozone to observed effects. While the SP concludes that there is not sufficient information yet to apply flux-based models, it indicates that flux-based models should be able to predict ozone injury and damage more accurately than currently-available concentration-response models. (page 7-20)

Seventh, the SP notes several findings from the flux-based experiments and model development that are pertinent to understanding and estimating ozone vegetation assessments in the current review. These include the findings that not all stomata uptake results in a reduction in yield and that there is a non-linear relation between ozone uptake and plant injury. (page 7-20) In addition, where there is a disconnect between the timing of peak exposure and maximal stomatal conductance, the predicted effect of the exposure is an over-estimation. (page 7-20)

Finally, the SP indicates that staff will consider whether and how additional flux/uptake-related factors could be combined with existing cumulative, peak weighted indices in order to develop an air quality index that is a better surrogate predictor of vegetation risk. (page 7-23) This is to be commended. However, no specifics are presented as to how that might be done. The second draft SP should provide specific suggestions or examples

for CASAC and public review. As Musselman et al.<sup>40</sup> have recommended, any flux-based index should take into account detoxification mechanisms.

### **Section 7.3 Vegetation impact assessment**

In the previous review, EPA staff carried out analyses similar to those discussed in this section of the draft SP. In the 1996 Federal Register notice proposing revised primary and secondary standards, the Administrator recognized many limitations and uncertainties in the available data and analyses. The proposal acknowledged that “the selection of a single averaging time for a national standard will of necessity be a compromise...” and “specifying the form of a seasonal exposure index to correspond to the relationship between vegetation response and ozone exposure is complicated by the many biological variables that influence the uptake of ozone by the plant and plant responses to such uptake.” (61 Federal Register, 65738, December 13, 1996)

The proposal also acknowledged that “because of a lack of monitoring data, national air quality typical of agricultural crop growing areas has not been characterized.” (61 FR at page 65740) Some of the most important caveats and uncertainties concerning the exposure and risk assessments for crop yield that were listed in the proposal included “extrapolating from exposure-response functions generated in open-top chambers to ambient conditions,” “the lack of a performance evaluation of the national air quality extrapolation,” “the methodology to adjust modeled air quality to reflect attainment of various alternative standard options,” and “inherent uncertainties in models to estimate economic values associated with attainment of alternative standards.” (61 FR at page 65740)

Discussing the quantitative analyses carried out at that time, the proposal indicated that “the Administrator acknowledges the significant uncertainties in these analyses and recognizes that these benefits should be regarded as rough approximations.” (61 FR at page 65743)

When the final standards were promulgated, the Administrator decided that based on the present limits of the scientific evidence of ozone effects and of rural air quality data, it was not appropriate to move forward with a seasonal secondary standard at that time. The decision was based in large part on the substantial uncertainties in the exposure, risk and valuation analyses and on the lack of air quality data in rural and remote areas. At that time, the Administrator reiterated her intention, that was expressed in the December 1996 proposal, to expand the rural ozone monitoring network (62 FR 38877-78 July 18, 1997)

Despite the Administrator’s stated intention to expand rural ozone monitoring, such an expansion has not taken place. Therefore, the continuing lack of sufficient ozone

---

<sup>40</sup> R. Musselman, A. Lefohn, W. Massmen, and R. Heath, “A critical review and analysis of the use of exposure- and flux-based ozone indices for predicting vegetation effects,” *Atmospheric Environment*, in-press, available on-line at [www.sciencedirect.com](http://www.sciencedirect.com).

monitoring in rural, remote, and agricultural areas remains a major impediment to the development of secondary ozone standard(s).

The SP acknowledges that ozone studies published since the 1996/1997 review haven't fundamentally altered the conclusions from the 1996 Criteria Document. The text raises the possibility, however, that new tools and refinements may allow for reduction in the uncertainties in the analyses. However, one element of the analysis that has not changed since the last review is the source of the crop loss C-R functions, the NCLAN (National Crop Loss Assessment Network) data that was obtained in the 1980's, and the tree-seedling C-R functions that were used in the 1996 review.

As discussed below, there are major issues and uncertainties related to the use of these C-R functions. Many of these were identified in the 1996 review and reiterated in the draft 2005 CD and the draft SP. In fact, the SP indicates that the planned updates to the risk assessments will still be limited by important data gaps and uncertainties. (page 7-2) A key issue or threshold question that should be addressed by CASAC and by the Agency is whether, without any substantial new C-R functions, the new assessment reduces the uncertainties in the risk estimates by any substantial amount. As the planned assessment is carried out and reported in the next draft SP, the staff should explicitly consider and discuss the extent to which the major sources of uncertainty identified during the 1997 rulemaking have been addressed.

### **Section 7.3.2 Air quality analyses**

The overview for this section correctly points out that it is a great challenge to characterize ozone air quality for the vast rural areas of the country that do not have ozone monitors. As noted above, an expansion of rural monitoring has not taken place since the last review. However, there have been improvements in photochemical grid modeling that may be of help. The EPA plan is to use existing ozone data and ozone outputs from the CMAQ model to generate ozone exposure surfaces in various regions for base year 2001 "as is" air quality. There are several key issues with the proposed approach that are not discussed sufficiently in the SP.

First, there is insufficient discussion of the performance of the CMAQ model and the uncertainty associated with its use. Although the overall CMAQ project has a very active performance evaluation program, only limited results are publicly available to date. The SP references a recent initial evaluation of the performance of CMAQ version 4.5 (Appel et al. 2005) for the eastern U. S. The prediction of daily max 8-hour ozone is described as fairly good but it is noted that CMAQ consistently over-predicted ozone at night. It is also noted that any estimates of ozone in complex terrain are very uncertain. However, it is argued that ozone gradients are low in flat, rural areas where important crops and vegetation grow and that night-time predictions are less an issue with the 12-hour SUM06 metric.

The Appel et al. 2005 evaluation is primarily focused on the CMAQ PM predictions and contains only limited analysis of the ozone predictions. Figure 15 below from the Appel

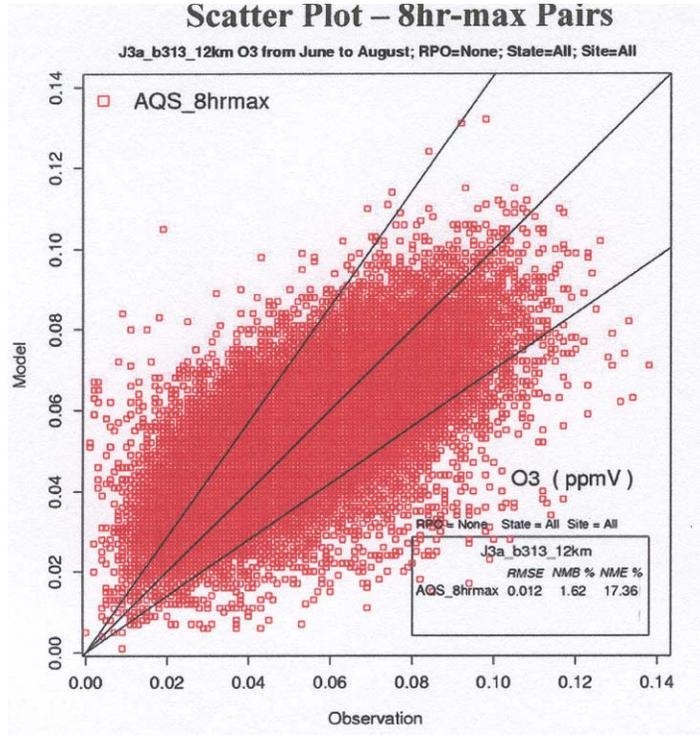
et al. evaluation is a scatter plot of 8-hour maximum predictions versus observations for the June through August 2001 period. The considerable scatter demonstrates a wide range of over- and under-predictions of the observations. For example, at the 0.06 ppm concentration level, which would be critical for a SUM06 standard, the model predictions range from 0.02 ppm to 0.09 ppm. In addition, there are numerous predictions of 0.06 ppm or higher on days with observations at 0.04 ppm or less. While the overall performance of CMAQ can be described as fairly good, the wide scatter results in a large number of false positives – where the model predicts concentrations greater than 0.06 or 0.08 ppm and the observations are below these thresholds – as well as a large number of false negatives – where the model predicts concentrations below 0.06 or 0.08 ppm and the observations are above these thresholds. The failure of the model to correctly predict the presence of elevated ozone a substantial portion of the time is troubling and is another form of uncertainty that needs to be considered by the Agency. A recent evaluation<sup>41</sup> of three other regional models also showed the inability of current regional model formulations to forecast exceedances, noting that only a small percentage (between 6 and 36 % depending on model and metric) of exceedances can be expected to be forecast correctly.

In addition, Appel et al. 2005 did not specifically evaluate the model performance in rural areas nor did it evaluate the model's predictions for ozone precursors. Appel does include several plots showing the model performance for specific locations, including two rural sites. Figure 16 below from Appel et al. 2005 shows the average (and 25<sup>th</sup> through 75<sup>th</sup> percentile) diurnal predictions and observations for the three-month period at two rural sites. The model performance at the Keeley Park, NC site is similar to the overall model performance for all sites – with a considerable over-prediction overnight and a much smaller diurnal variation than present in the observations. However, the model performance at the Leslie, GA site (which is about 100 miles south of Atlanta) is troubling. The model shows little diurnal variation for the three-month period resulting in over-predictions at night and under-predictions during the normal peak hours. Since no careful evaluation of the rural predictions of CMAQ has been reported, and it is not clear how prevalent these two patterns are, these differences raise important questions as to the utility of CMAQ either for generating ozone exposure surfaces or for predicting the impact of emission controls in rural areas.

The SP acknowledges that model predictions in complex terrain are very uncertain and the Appel et al. 2005 evaluation indicates that the performance in the 36 km grids deteriorated compared to that in the 12 km grids. These findings raise additional questions concerning the use of CMAQ or any model for predicting ozone in rural locations across the country, and particularly in the western half of the country.

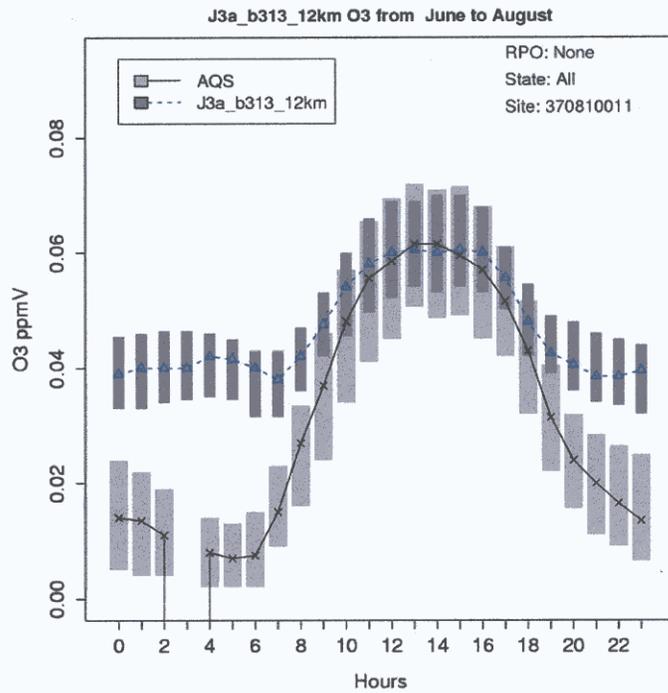
---

<sup>41</sup> D. Kang et al., *J. Air & Waste Manage. Assoc.*, **55**, pages 1782-1796 (2005)



**Figure 15 – From Appel et al. 2005**

### Rural Site (Keely Park, NC)



### Rural Site (Leslie, GA)

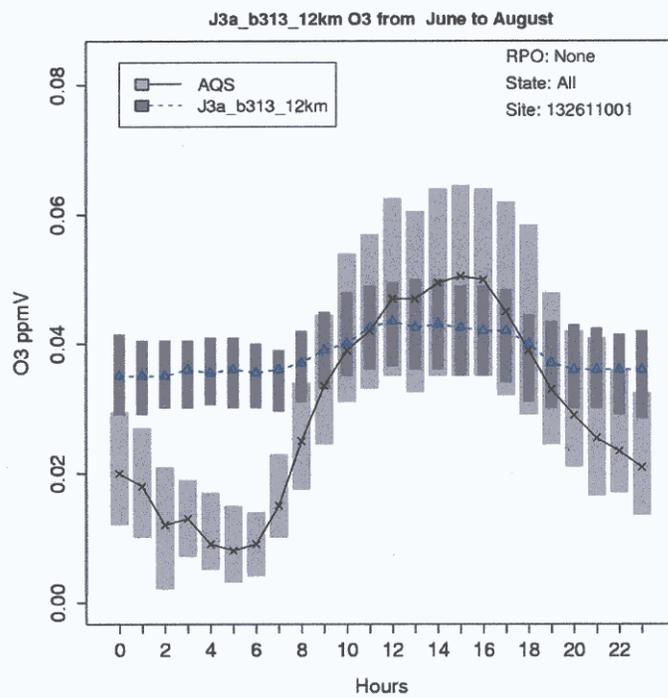


Figure 16 – From Appel et al. 2005

There are additional studies evaluating the current version of CMAQ or its predecessors that are relevant and should be considered in the SP. In a series of papers evaluating the performance of regional modeling systems for ozone, ozone precursors and meteorological parameters, Holgrefe et al.<sup>42</sup> reported that the model performance was good on the synoptic and seasonal time scales but deteriorated on the diurnal and intraday time scales. They point out that the lack of agreement between meteorological model outputs and observations on the intraday and diurnal time scales has important implications to air quality models that use these meteorological input fields. In particular, ozone changes cannot be simulated accurately when the meteorological drivers are not simulated accurately. For ozone, the intraday components and the amplitude of the diurnal component was poorly captured. They point out that the time and space scales needed for exposure assessment and forecasting were not captured by the modeling systems they evaluated, which include one of the predecessors of CMAQ.

For ozone precursors, the traditional statistical measures of performance - mean normalized bias and mean normalized gross error – were much higher than for the ozone predictions. While there are no EPA guidelines for evaluating model performance for precursor concentrations, the fact that precursor concentrations are very poorly modeled should give one great pause in using regional models. If the precursor concentrations in the model are wrong, even if the model predicts the correct ozone, it is getting the ozone right for the wrong reasons and there are other compensating errors in the model.

Each fall, EPA and NOAA hold a CMAS modeling conference. There are several papers from the September 2005 conference that are available as extended abstracts or presentations on the CMAS website that provide additional insight into CMAQ performance. Liu et al.<sup>43</sup> reported on the application of CMAQ in the southeast U. S. They used process analysis to provide insights into how the model predictions are obtained. They report that vertical transport and dry deposition are the two major contributors to the changes in ozone mixing ratios. In addition, Yu et al.<sup>44</sup> reported on an evaluation of CMAQ in which detailed composition and meteorological measurements at several ground sites and from aircraft were compared to CMAQ predictions. They report poor performance for precursors and note that, for three sites where ground level measurements of the rate of photolysis of nitrogen dioxide were made, the model reproduces the observations within a factor of 1.5 only about half the time. They also point out that a 50 % uncertainty in the photolysis rate can result in up to a 40 ppb uncertainty in ozone predictions.

---

<sup>42</sup> C. Holgrefe et al., *Atmospheric Environment*, **35**, pages 4159-4174 (2001); C. Holgrefe et al., *Atmospheric Environment*, **35**, pages 4175-4188 (2001); J. Biswas, C. Holgrefe et al., *Atmospheric Environment*, **35**, pages 6129-6149 (2001).

<sup>43</sup> C. Liu et al., "Simulating ozone and PM with CMAQ in southeast U. S.: Application, evaluation, and process analysis," extended abstract for presentation at 4<sup>th</sup> Annual CMAS Models 3 User's Conference, September 26-28, 2005, Chapel Hill, NC.

<sup>44</sup> S. Yu et al., "A comprehensive evaluation of the Eta-CMAQ forecast model performance for ozone, its related precursors, and meteorological parameters during the 2004 ICARTT study," presented at the 4<sup>th</sup> Annual CMAS Models 3 User's Conference, September 26-28, 2005, Chapel Hill, NC.

When these various evaluations of the CMAQ model are considered together, they demonstrate a number of problems with the model that limit its usefulness at this point in time. The poor performance at night clearly demonstrates problems with nighttime removal processes. These could be issues related to vertical mixing, dry deposition, and/or nighttime chemistry. There may well be problems with the vertical mixing and dry deposition during the day and, since these parameters are key determinants of daytime ozone, there may be compensating errors. Among the meteorological parameters, wind speed has the poorest model predictions. However, precursor concentrations are determined primarily by wind speed and vertical mixing. Thus, errors in these parameters will translate into errors in precursor concentrations that are very poorly simulated in current regional models. Errors in the precursor concentrations translate into errors in local chemical formation of ozone. The poor performance for photolysis rates is an indication of an inability to properly simulate the effect of clouds and also translates into errors in the local chemical formation of ozone. Taken together, all these issues with current regional or national models lead to greater uncertainty than currently portrayed in the SP. The second draft SP should more thoroughly discuss these model performance issues.

Second, the specification of grid size in CMAQ is critical to adequate model performance. Some of the uncertainties noted above are directly related to the spatial resolution of the model. It has been shown in numerous studies that grid resolution affects the magnitude of ozone formation and loss processes as well as transport.<sup>45</sup> Since ozone production is non-linear and highly sensitive to precursor ratios, large horizontal grids that mix emissions in the model in ways that do not occur in the real world can provide misleading results. In an evaluation of CMAQ results for the simulation of a high ozone episode in Houston, it was found that high horizontal resolution (1-km) was needed to adequately simulate peak ozone levels. Since the transport of high man-made ozone into downwind rural areas is one of the key issues involved in ozone vegetation effects, the coarse grid resolution in the continental simulations in the planned assessment may lead to substantial uncertainty and/or bias.

Third, the interpolation module in the BenMAP doesn't take chemical behavior into account. This is one of the reasons why the staff is interested in including CMAQ outputs as well as monitor data. However, there are large areas of the country with little or no monitoring data or verified emissions inventories. Thus, the SP indicates that staff is considering several options. Until the CMAQ results are verified or there is a major increase in rural ozone monitoring, there will be considerable uncertainty in the exposure surfaces generated with any of the approaches discussed. This component of uncertainty was identified in the 1996 review as "the lack of a performance evaluation of the national air quality extrapolation."

An inspection of the previous estimates of exposure surfaces reveals several additional concerns. For example, the "as is" estimate of SUM06 for the year 1990 is shown in

---

<sup>45</sup> See Cohan et al., *Atmospheric Environment*, **40**, pages 126-135 (2006); Liang and Jacobson, *Atmospheric Environment*, **34**, pages 2975-2982 (2000); Chock et al., *Atmospheric Environment*, **36**, pages 4649-4662 (2002); Gillani and Pleim, *Atmospheric Environment*, **30**, pages 2043-2059 (1996).

Figure VII-10 of the 1996 SP (the comparable map for W126 is on page E-1 of Appendix E). While there are high SUM06 values estimated in and around many major cities, the overall pattern does not fully mimic the areas of high anthropogenic emissions. The substantial SUM06 and W126 values throughout the Southwest and the Southeast suggest that SUM06 is influenced by both background ozone and natural emissions as well as by anthropogenic emissions. For comparison, the spatial estimates from “as is” ozone in 2001 are given in the maps in Figures AX3-5 to AX3-19 of the January 2005 draft CD. In this case, maps of both the mean estimates and the 95% confidence intervals are presented. When the maps of the mean estimates for 1990 and 2001 are compared, there are many similarities, but there is one substantial difference. The methodology used for the 2001 case appears to spread the influence of major source areas much more than the methodology used for the 1990 case. Moreover, the 95% confidence intervals for the 2001 estimates are extremely wide in the central and western U. S. Thus, the uncertainty in the estimates is quite large and both seasonal indices appear to include contributions to some extent from background ozone.

As shown above in comments on Chapter 2, the maps of observed SUM06 for recent years (Figures 5 to 7) show wide variations from year to year. There is also a general North to South gradient and the overall pattern suggests both a contribution from U. S. man-made emissions and a background component.

In addition to generating “as is” ozone air quality, EPA plans on generating ozone surfaces meeting the current standard and meeting (as yet unspecified) alternative standards. There is another set of key issues associated with this portion of the plan. First, and most importantly, there is no indication that the chemistry of ozone formation will be used in this portion of the Plan. The various forms of rollback in BenMAP are all arbitrary. In contrast to an arbitrary rollback of ozone concentrations, efforts to attain any alternative standard in the real world will involve combinations of national, regional, and local controls that will change the distribution of ozone concentrations at different locations in ways that can only be determined with photochemical models. Until such modeling is carried out, there will be continued uncertainty associated with using any form of arbitrary rollback. This component of uncertainty was identified in the 1996 review as uncertainty due to “the methodology to adjust modeled air quality to reflect attainment of various alternative standard options.”

The evaluation of the quadratic method to analyze differences between high and low ozone years noted on page 7-34 is evaluating the impact of meteorology rather than the impact of emission changes. Therefore, it is not a sufficient test. The planned analysis of the trends at the rural Crestline monitor in the San Bernardino Mountains, while it does address both emissions and meteorology issues, is likely location-specific. The site is immediately downwind of a very large source area known to have unique meteorological and topographical characteristics that affect ozone formation and transport.

Second, although Chapter 2 indicates that estimates of background used in the SP are based on the Fiore et al. 2003 modeling study, Chapter 7 is silent on exactly how the background of uncontrollable ozone will be treated in the analysis. Previous AIR

comments<sup>46</sup> and the discussion above for Chapter 2 raise several issues regarding the CD and SP's discussion of policy relevant background and the Fiore et al. study. These include:

- The definition of policy relevant background (PRB) in the draft CD (and now the SP) is wrong; it omits consideration of the emissions from agricultural activities and omits consideration of the contribution to ozone from anthropogenic emissions in Mexico and Canada.
- In order to evaluate the range and variability in PRB, the CD and SP need to consider both observations and models.
- There are several issues with the one modeling study that the SP relies on to estimate PRB making its use problematic. The second draft CD acknowledges many of these issues and recommends using a suite of models including models with greater resolution to estimate background.
- The widespread nature of yearly maxima 8-hour ozone concentrations of 0.06, 0.07 and higher at Western national parks and monuments and in rural areas that are distant from large urban areas indicates the presence of a substantial regional policy relevant background of ozone.

Because of the continuing uncertainty as to what portion of the ozone in rural and remote areas is controllable through control of man-made U. S. emissions, there will be uncertainty related to any of the possible choices for defining or treating policy relevant background in the SP. The SP should provide a realistic range of estimates for a properly defined PRB for each of the vegetation exposure metrics considered in the SP.

Third, it is not clear which exposure metrics will be evaluated. The SP indicates that, at a minimum, a 12-hour, 3-month SUM06 index and an 8-hour average index will be evaluated. However, it is also indicated that additional indices may be selected. During the last review, the sigmoid-weighted W126 index was considered in detail. Since this index was selected by the Federal land managers for protecting vegetation in Class 1 areas (see draft CD at page 9-194) and since it avoids the unrealistic assumption of an arbitrary cut-off value, it should be included in the analysis.

There are also questions as to how the 8-hour average index will be defined. Since the current 8-hour primary and secondary standards are defined as the average over three years of the 4th highest 8-hour average in a year, defining an equivalent 8-hour metric for

---

<sup>46</sup> J. M. Heuss, Comments on the First External Review Draft of "Air Quality Criteria for Ozone and Related Photochemical Oxidants," EPA 600/R-05//004bA, January 2005, Prepared for General Motors Corporation, Air Improvement Resource, Inc., April 26, 2005; J. M. Heuss, Comments on the Second External Review Draft of "Air Quality Criteria for Ozone and Related Photochemical Oxidants," EPA 600/R-05//004aB, August 2005, Prepared for General Motors Corporation, Air Improvement Resource, Inc., November 30, 2005.

an annual growing season and for the NCLAN experiments will involve some thought and analysis.

### **Section 7.3.3 Crop risk/benefits assessments**

The overview points out that staff will be updating the previous review's analyses but still using the NCLAN crop loss C-R functions. This means that the issues and uncertainties related to the NCLAN experiments that were identified during the previous review are still relevant for the current review.

The NCLAN experiments were carried out in the 1980s and the study design was a major improvement over earlier approaches. The study was designed to provide input for an estimate of the economic effects of ozone on U. S. agriculture for the review of the ozone standard. Nevertheless, there were important issues concerning the NCLAN studies raised by the scientific community and acknowledged by the Agency during the previous review. For example, the 1996 Staff Paper indicated that some plant scientists continue to express concern that NCLAN was "designed and conducted in a way that results in overestimation of ozone impacts." (1996 Staff Paper at page 194) Three issues that were acknowledged by the Agency at the time are still relevant. First, the 1996 SP indicated:

"...the modified ambient treatments contained numerous high peaks (ozone concentrations equal to or above 0.10 ppm), occurring more frequently than in typical ambient air quality distributions. Such exposure patterns have raised questions among some researchers as to how much of the plant's response was a result of having an excessive number of high peaks versus a cumulation of more moderate exposures." (1996 Staff Paper at page 194)

The question of the relative role of peaks and mid-level exposures in ozone vegetation effects is still a major issue and concern in the current review.

Second, the 1996 SP indicated:

"...the charcoal filtered chambers used to establish baseline crop yield loss were exposed to approximately 0.025 ppm ozone, which is lower than the range of 0.03 to 0.05 ppm identified in Chapter 4 of the staff paper as the value for seasonal background ozone levels. The result of using this lower level of 0.025 ppm is an overestimation of yield loss relative to that expected using 0.03 to 0.05 ppm." (1996 Staff Paper at page 194)

The question of the appropriate control or background for the yield loss estimates is still a major issue and concern in the current review. For example, the January 2005 draft CD notes that the issue of the background ozone level is important to all estimates of vegetative damages due to ozone, noting that recent research by Lefohn et al. 2001 have suggested that background levels may be considerably higher than assumed in some of the previous economic assessments. The CD goes on to acknowledge that, if that is

true, “then the economic damages estimated in studies with lower background levels will be overstated.” (January 2005 draft CD at page 9-327)

Third, the 1996 SP indicated:

“Review of the NCLAN data indicates that differences in ozone sensitivity within species may be as great as differences between species with substantial variation in sensitivity from year to year.” (1996 Staff Paper at page 195)

This finding raises two issues for the current review. The first is the variation in sensitivity from year to year. It is reflective of the well known fact that environmental variables affect gas exchange and ozone uptake. The SP notes the growing body of information on ozone flux and vegetation damage, but indicates that it is not sufficient at this time to use in a quantitative way. Manning has argued that any cumulative value air quality standard, such as SUM06, that does not take into account the variables that affect ozone uptake via stomata has no biological basis or relevance, and can lead to overestimation of ozone injury on plants in the field.<sup>47</sup> As Manning notes, SUM06 is based on two assumptions --- that plant exposure to ozone equals plant response, and that there is an identifiable uniform threshold value for plant injury caused by ozone. Both these assumptions are known to be incorrect. The SP should acknowledge and discuss the uncertainty associated with these simplifying but incorrect assumptions.

The finding of as great a difference within species as between species raises the issue of the relevance of the cultivars that were used in the NCLAN study to the cultivars that are being planted today. This is acknowledged on page 7-35 of the SP. However, the text indicates that, in general, new crop varieties are not specifically bred for ozone tolerance. In addition, the text indicates that the fact that ozone levels are not consistent from year to year does not allow crop breeders to select for ozone tolerance under natural conditions. This second point is not true. Since ozone in varying degrees is always present in natural conditions, there will always be a tendency for ozone resistant cultivars to be selected. Since the SP indicates that 2001 county level crop planting data will be obtained to create maps of the planting data for each species/cultivar of commodity crop, the Agency will be in a position to evaluate the relevance of the NCLAN cultivars.

Another issue that was raised during the review of the January 2005 draft CD and the October 2005 CASAC teleconference and not addressed properly in the SP is the concern that the ozone concentrations at the “standard measurement height” may not be the same as the ozone concentration at plant height is noted on page 7-27 and acknowledged as a bias on page 7-30. However, it is dismissed as a small bias on page 7-30 referring to data indicating a 7 % effect. Since this effect will result in an overestimate of the ozone exposure of the crop plants and the predicted crop loss (See CASAC June 22, 2005 letter to the Administrator at page C-57 and Grunhage et al. 1999 reference in Chapter 9 of the CD), the assessment should correct for this effect. A 7 % reduction in total ozone will translate into a much larger percent change in man-made ozone or SUM06. At a

---

<sup>47</sup> W. J. Manning, *Environmental Pollution*, **126**, pages 375-379 (2003)

minimum, a correction factor should be applied to the ozone data to determine the magnitude of the bias in the resulting crop loss estimate.

### **Section 7.3.3.3 Economic benefits assessment**

Since the output of the Agriculture Simulation Model (AGSIM) or any other model used to provide estimates of the benefits of ozone control will be important factors in the ultimate decision regarding secondary standards, the uncertainty in the estimate should be carefully evaluated and reported. Both uncertainties due to the model formulation and uncertainties due to the model inputs should be evaluated and discussed in the Staff Paper. In particular, the sensitivity of the results to different background assumptions should be explicitly evaluated.

The January 2005 draft CD noted the need for improving our understanding of the temporal and spatial characteristics of ozone as it relates to the implications for crop yields, production, and producer profits. (Draft CD at page 9-327) For example, the CD pointed out that most economic studies are static (comparing economic activity at one ozone level versus that at an alternative level) whereas the changes in air quality will be gradual which means that producer responses will be gradual, rather than abrupt. The implications of the simplifications and assumptions in the economic model for the overall assessment need to be discussed in the Staff Paper. In the previous review, the Administrator acknowledged that the overall benefits estimates are rough approximations. Given the lack of new or biologically-relevant C-R data, it is not clear whether new benefits estimates will be substantially more accurate or precise than the previous estimates.

### **Section 7.3.4 Tree risk assessments**

As is the case with crops, there is little new data upon which to draw. The proposed analysis of ozone effects on Ponderosa Pine in the San Bernardino Mountains is not particularly relevant to the rest of the U. S. or the setting of a national standard. The area, being downwind of Los Angeles, has been subject to ozone levels far in excess of either federal or California ozone standards for many decades. As noted in the SP, there are many uncertainties and limitations in the modeling framework being considered for tree growth and economics. As such, the exercise is more appropriate for a research project than for application to standard setting.

The analysis and comparison of the incidence and severity of foliar symptom data from the Forest Service bio-monitoring sites with ozone exposure estimates is a useful project. However, it should not be limited to comparisons with the exposure surfaces. It would also be useful to compare the data separately with the CMAQ ozone output and with ozone surfaces generated with just interpolated observations. Since the data is on foliar injury to plants known to be sensitive to ozone and the SP acknowledges that there is not always a direct relationship between visible foliar symptoms and growth (page 7-46), it is not clear how this data will be used in the standard-setting process.

### **Summary thoughts on Chapter 7**

The results of the various vegetation risk estimates will not be available until the second draft SP is issued in April 2006. The draft SP indicates that staff will identify sources of uncertainty and treat them qualitatively and quantitatively in the second draft SP. (page 7-27) Because the issue of uncertainty loomed so large in the previous review, the Agency should use sensitivity analyses wherever possible to bound the estimates. While it may not be possible to do a formal probabilistic risk assessment that incorporates all of the uncertainties in the risk assessment during this review, the Agency should lay the groundwork for such an analysis during the next review.

If staff and/or CASAC is going to recommend a seasonal secondary ozone standard, the question of whether tools are available to implement such a standard becomes important. Therefore, a careful review of the capability of CMAQ or other possible models to simulate rural ozone, in general, and any seasonal ozone metrics under consideration should be carried out and included in the second draft SP. Given the issues of model performance outlined above and the inherent limitations of coarse grid models being unable to deal properly with sub-grid scale processes, it is difficult to see how a seasonal secondary ozone standard could be effectively implemented.