Comments on Final Ozone Staff Paper and CASAC Recommendations for Primary and Secondary Ozone Standards

Jon M. Heuss
Air Improvement Resource, Inc.

Prepared for the
Alliance of Automobile Manufacturers

March 1, 2007

Executive Summary

The U. S. Environmental Protection Agency (EPA) final Ozone Staff Paper (SP) and the Clean Air Scientific Advisory Committee's (CASAC) October 24, 2006 letter to the Administrator makes recommendations regarding the primary and secondary standards that are based on insufficient information. Further analysis of the epidemiology, toxicology and background will show that the science does not support lowering the ozone National Ambient Air Quality Standards (NAAQS).

While both EPA staff and the scientific advisory group recommend extreme value statistical forms for the standards, the EPA analysis did not evaluate the extreme values of the background of ozone that cannot be controlled through reduction of U. S. man-made emissions. As a result, the ranges recommended by both groups overlap with the extremes of background. In previous reviews, both means and extremes of background were evaluated. The Criteria Document (CD) indicates that the global model EPA relied on in the current review is unable to simulate local maxima. The large grids in the global model mix air parcels of different composition and origin in an unrealistic way, severely limiting its ability to accurately simulate local photochemical activity, intercontinental transport, and the impact of stratospheric air inserted into the troposphere.

The risk of the respiratory effects identified in controlled human studies is substantially reduced from that estimated in the previous review. The dose-response is now known to be non-linear and studies now 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 these effects are transient, of a reflex nature, and not life threatening.

There are four ways in which the risk assessment of lung function effects markedly overestimates the number of incidences of concern. First, the impact of just meeting alternative standards is estimated by rolling back ozone only at monitoring sites that exceed the applicable standard. In reality, national and local emission controls reduce ozone throughout a metropolitan area not just at the monitors that exceed a given

standard, resulting in peak ambient levels below the standard at all monitors except the design value monitor. Second, no correction was made for the fact that ozone exposure is lower at “person” height compared to “measurement” height. This effect was corrected in the vegetation risk assessment but not in the human risk assessment. Third, the approximation that all exposures between 13 and 27 EVR (equivalent ventilation rate) respond as though they were at 20 EVR overestimates the risk substantially. Fourth, values produced by the ventilation rate algorithm exhibit an excessive degree of variability overstating the number of exposures associated with strenuous work or play. When predicted ventilation rates were compared with measured rates in a database of over 30,000 person-days from a cohort of over 2200 free-living individuals between the ages of 3 and 96, the model overpredicted mean ventilation rates for persons below age 11 and over age 40 and the model had a much higher standard deviation at all ages.

EPA and CASAC continue to overestimate the magnitude and consistency and underestimate the uncertainty in the results of acute epidemiologic studies, especially for emergency room visits and hospital admissions for asthma and other respiratory disorders, school absences, and mortality. Although there clearly are positive associations with ozone in the literature for many health endpoints, there are also positive associations for the same health endpoints for many other pollutants. In addition, in systematic studies the overall pattern of associations is similar for each of the pollutants and the overall range of associations is not biologically plausible. While there is now a much larger set of studies than was available in 1996/7, the implausibly wide range of results from positive to negative in systematic analyses, and the newly recognized issues of model selection and publication bias cloud the interpretation of the data as causal. EPA overstates the likelihood of causality, particularly for the more serious endpoints.

With regard to the secondary standard:

- The extremes of background confound the ranges recommended for the secondary standard.

- The analysis used to discriminate between cumulative standards and the current 8-hour standard is flawed because the rollback reduces ozone only at sites that do not attain the 8-hour standard. Therefore the need for a cumulative standard is overstated.

- The Europeans have used cumulative standards only as a first step toward establishing a more biologically-relevant flux-based standard.

- The estimates of crop loss from ozone are substantially lower than those of the previous review.

- The recommended range for a cumulative standard needs to be adjusted upward to account for the reduction in ozone at plant height compared to measurement height.

- The question of whether monitoring and modeling tools are available to implement such a standard needs to be addressed by EPA and CASAC.

Introduction

The U. S. Environmental Protection Agency issued the final Ozone Staff Paper[1] along with Technical Support Documents on January 31, 2007. The Clean Air Scientific Advisory Committee is holding a teleconference on March 5, 2007 to consider providing additional comments on the final SP to EPA. In particular, CASAC indicated that it will be reviewing Chapters 6 and 8 regarding recommendations for the primary and secondary standards.[2] Although CASAC had recommended a primary 8-hour ozone standard in the range from 0.060 to 0.070 ppm (with a form similar to the existing 0.08 ppm standard), the final SP recommends considering a standard within the range of somewhat below 0.080 ppm to 0.060 ppm. For the secondary standard, EPA staff concurred with the CASAC recommendation for a cumulative form that extends over an entire growing season, proposing a three-month seasonal W126 standard between 7 and 21 ppm-hr. The upper end of the range, 21 ppm-hr, however, was chosen to be equivalent to the SUM06 standard of 25 ppm-hr that was considered but not adopted in 1997 rather than the 17 ppm-hr level recommended by CASAC.

In the following, Section I provides comments on the primary standard. Section II provides comments on the secondary standard. Section III provides comments on the role of CASAC. Appendix A provides additional detail regarding uncontrollable background and Appendix B provides additional detail regarding the interpretation of epidemiological studies.

Section I Comments Regarding the Primary Standard

Background or Uncontrollable Ozone

There are important points that EPA and CASAC should consider regarding the implications of uncontrollable background ozone for standard setting. First, there should be a detailed evaluation of uncontrollable background expressed in the same form as the metric for the proposed standard. Second, estimates of background should be based on the full range of models and observations that can be brought to bear on the issue instead of just one global model. The lack of consideration of the extremes of background is an omission that led both the SP and CASAC to assume that the standards recommended are attainable.

The SP and CASAC do not include information regarding the extremes of uncontrollable or policy-relevant background, the ozone that is not subject to

[1] U. S. Environmental Protection Agency, “Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information: OAQPS Staff Paper,” EPA-452/D-07-003, January 2007.
[2] R. Henderson, CASAC Peer Review of the Agency’s 2nd Draft Ozone Staff Paper, EPA-CASAC-07-001, October 24, 2006.

control by reduction of man-made U. S. emissions. As a result, the range recommended in the SP overlaps with the extremes of background.

Uncontrollable background ozone should be carefully evaluated and expressed in the same form as the metrics being considered for the standard. By focusing on the mean background rather than the background expressed on the same statistical form as that of the standard, neither EPA nor CASAC has adequately evaluated whether the ranges they recommend can be achieved even with complete elimination of man-made emissions that might be controlled.

The lower end of the range recommended by CASAC, 0.060 ppm, would put virtually the entire country out of attainment including sites that are far removed from any significant man-made emissions. Figure 1 shows that only a few low-level coastal western sites (that are influenced primarily by marine air low in ozone) and a few of the western sites in the northern-most states would attain a standard at the low end of the range recommended by CASAC. In addition, many western sites removed from significant man-made emissions have ozone design values in the range of 0.069 to 0.075 and would not attain even the upper end of the CASAC-recommended range. Furthermore, the upper end of the range recommended by CASAC, 0.070 ppm, was specifically rejected by the Administrator and CASAC in the previous review as being too close to background. During previous reviews of the ozone standards, both means and extremes of background were evaluated and considered.

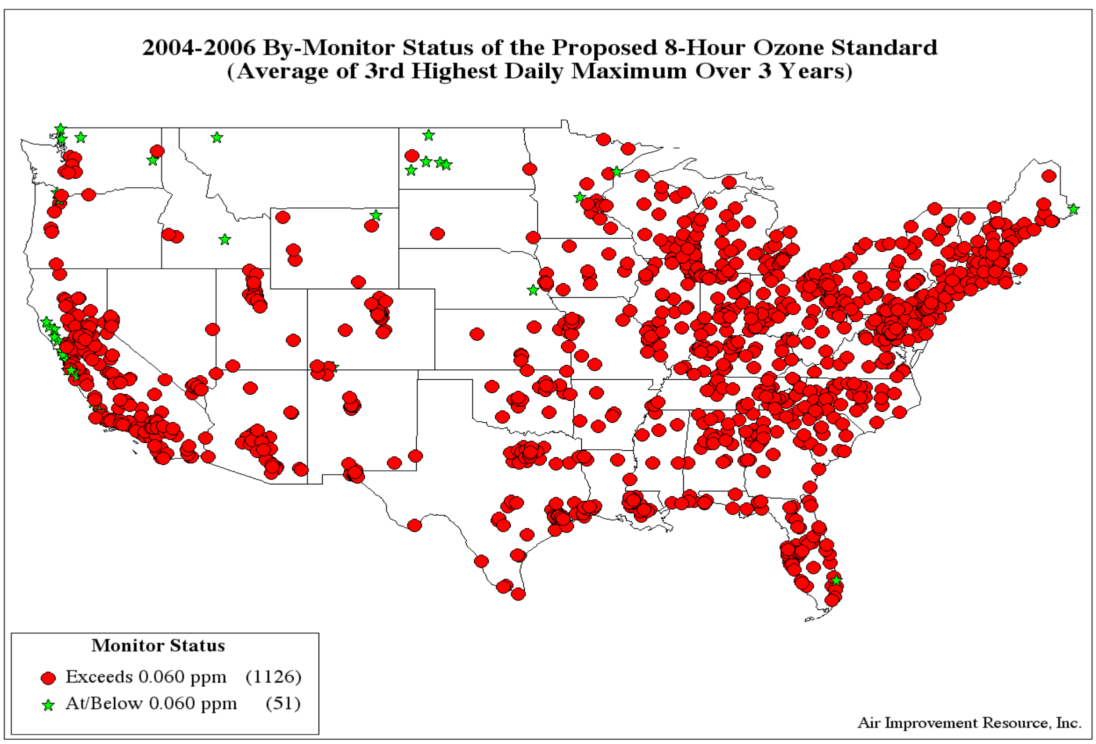


Figure 1

The global model EPA relies on to estimate background is not capable of resolving the phenomena that contribute to attainment or non-attainment of an extreme value standard throughout the country.

The SP relies solely on the GEOS-CHEM global model (Fiore et al. 2002, 2003) to estimate mean background. However, the ozone Criteria Document (CD) notes many important limitations of global models including the GEOS-CHEM model and the need for additional work to evaluate policy-relevant background with an ensemble of models comparing the model results with observations. [3] The CD notes the need to include models with greater resolution to understand variability on shorter time scales and variability due to processes that are not captured in the global models. Until this is done, a range of background estimates from models and observations should be used in the ozone review to estimate the distribution of background, with an emphasis on the extremes.

The coarse horizontal resolution is a particularly severe limitation. Since ozone formation is a highly non-linear process and mixing is a linear process, the smearing of emissions into large grids results in precursor concentration fields that are known to be incorrect. Acknowledged problems with the treatment of clouds in global models also means that local temperatures and UV flux are not accurately simulated. Even if a global model predicted the correct ozone concentration field it would be doing so for the wrong reasons. A global model would not be considered for use in developing a State Implementation Plan for either the current or a revised standard. Even the higher resolution Community Multi-scale Air Quality modeling system (CMAQ) performed extremely poorly in predicting SUM06 or W126 in the western U. S. as documented in Appendix A and in the secondary standard Technical Support Document (TSD).[4] For example, in the Western U. S. minus California, CMAQ had a normalized mean error and a normalized mean bias of over 400 %. One of the implications of this failure is that there are sources and phenomena that produce ozone concentrations of 0.06 ppm and higher in rural and remote locations that are not captured in the present global or regional models. Additional documentation of these concerns is provided in Appendix A.

Both models and observations need to be considered in estimating the distribution of background concentrations.

Because of the acknowledged limitations of global models, observations must be carefully evaluated for estimates of extreme values of background at specific locations. The CD notes the several processes that contribute elevated ozone in rural and remote locations, some of which are related to U. S. man-made emissions and some of which are related to uncontrollable background. While this makes total reliance on observations

---

[3] [3] U. S. Environmental Protection Agency, Air Quality Criteria for Ozone and Related Photochemical Oxidants, EPA 600/R-05/004aF, February 2006at page 3-55.
[4] J. Lehrer et al., Techical Report on Ozone Exposure, Risk, and Impact Assessment for Vegetation, EPA 452/R-07-002, January 2007;at Appendix A.

problematic, the severe limitations of global models for predicting extreme ozone events make reliance solely on global models (much less one specific global model) even more problematic. Therefore, both models and observations need to be used to determine the extremes of uncontrollable ozone before an ozone standard is proposed that may well be unattainable.

The evidence considered by EPA staff to support the contention that a 0.06 ppm standard is even theoretically attainable is the Fiore et al. 2003 modeling study that concluded that background never exceeds 0.04 ppm. However, the Fiore study acknowledges that “sub-grid scale local peaks are underestimated.”

In addition, the definition of policy relevant background used by EPA is overly restrictive. It assumes that all man-made Canadian, Mexican and all North American agricultural emissions can be fully eliminated. These are unrealistic assumptions that lead to an underestimation of uncontrollable background ozone in the U. S.

Controlled Human Studies

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 is substantially lower than originally thought in 1997 when the current 8-hour ozone standard was set.

Since the previous review, a number of findings and factors indicate that there is less risk of lung function decrements and symptoms and that the public health significance of those effects is substantially lower than originally thought in 1997. These include:

- The data extending lung function effects to lower concentrations establishes that the response is non-linear in contrast to the linear assumption made in the previous review

- Although there are many differences between the 1997 and 2007 risk estimate methodologies, it is clear from the second draft SP that the percent of the population undergoing exposures of concern is significantly smaller in the current modeling than originally thought in 1996/97[5].

- Although respiratory symptoms in healthy children were estimated in the 1996 review, that endpoint was not included in the current review since a number of field studies failed to find such effects.

- The CD now acknowledges that a neurally-mediated mechanism is responsible for the lung function decrements and symptoms seen in clinical studies. A body of studies now shows that the first effects of ozone are caused by reflex responses of vagal nerve receptors that involuntarily inhibit maximal inspiratory effort. This response is

[5] Second draft Staff Paper at page 5-71.

the body's reflexive reaction to the presence of an irritant gas and is unrelated to sensations of discomfort. The CD acknowledges that these effects are transient, of a reflex nature, and not life threatening.

- In the previous review, there was a consistency of CASAC and staff opinion that single, acute moderate lung function responses in healthy individuals should not be considered adverse. However, the results of the risk analysis in the current review focus on the number of subjects with one or more responses in a year.

- For asthmatics, staff previously concluded, based on discussions with medical experts, that single ozone exposures that resulted in moderate responses were not likely to interfere with normal activity nor to result in increased frequency of medication or the use of additional medications.

- While there was concern over the reversible short-term effects as precursors of irreversible long-term changes in the previous review, there are now several studies that found no long-term lung function changes in children that could be ascribed to ozone in the high ozone areas of California.

There are a number of reasons to conclude that the risk of respiratory effects is overestimated in the APEX model risk assessment

There are concerns that the APEX model and the resulting risk assessment overestimates the number of occurrences of elevated 8-hour ozone exposures during strenuous work or play in four ways.

1. The rollback approach underestimates the ozone improvement and risk reduction that will occur in conjunction with meeting a given standard. The approach estimates the impact of just meeting various alternative standards by rolling back ozone only at monitoring sites that exceed the applicable standard. In reality, a combination of national and local emission controls reduce ozone throughout a metropolitan area not just at the monitors that exceed any given standard, resulting in lower ambient levels than the standard at all monitors except the design value monitor.

2. The CD acknowledges that ozone exposure is lower at "person" height compared to "measurement" height. This effect was corrected in the vegetation risk assessment but not in the human risk assessment. In the vegetation risk, SUM06 was halved with a 10 % vertical correction. By analogy, a vertical correction in the human risk would halve the number of human exposures of concern at ground level.

3. The upper tail of the distribution of equivalent ventilation rates (EVR) is particularly important because it a critical factor in determining the frequency of exposures of concern. Previous AIR comments noted the approximation that all exposures between 13 and 27 EVR respond as though they were at 20 EVR substantially overestimates the risk since there are many more 8-hour occurrences of ozone

exposures at 13 EVR than at 20 and many more at 20 than at 27. This results in overestimated exposures in the upper tail of the distribution.

4. Previous comments also stated the importance of evaluating how the model's predictions of the number of elevated ventilation rate occurrences compare with real world data. This is a particularly important comparison that was not available to CASAC before the Panel provided its recommendations. Last year, Langstaff acknowledged the "values produced by the ventilation rate algorithm may exhibit an excessive degree of variability."[6] The final sensitivity analysis for APEX includes a comparison of predicted ventilation rates with mean values in the literature, but the upper tails of the distribution which impact the risk estimates were not compared.[7] In the comparison of the APEX modeled values with the measured ventilation rates from Brochu et al. 2006,[8] the model overpredicted mean ventilation rates for persons below age 11 and over age 40. More importantly, the model had a much higher standard deviation at all ages. This suggests that the upper percentiles of ventilation rate in the model are substantially above those measured in a database of over 30,000 person-days from a cohort of over 2200 free-living individuals between the ages of 3 and 96. The EPA should also rigorously compare the upper strata of the model with the upper strata of measurements. This is an important oversight because the upper percentiles of ventilation rate are responsible for the exposures that cause the perceived risk.

Even with significant over-estimates of the number of exposures associated with exercise, the APEX results presented in the SP demonstrate that the current standard limits the number of persons and person-occurrences of ozone exposures to 0.08, 0.07, and 0.06 ppm for 8-hours with moderate exertion (strenuous work or play). There are several ways to put the risk assessment findings in perspective. For example, one can focus on the fact that a given standard protects over 90 percent of adults or children all the time and the other few percent over 99 percent of the time. Or one can focus on the fact that there are a few percent of adults or children have one or more exposures of concern in a year. The SP focuses on the numbers of those exposed at least once per year and how that changes with various alternatives.

In the previous review, CASAC viewed the risk assessment (for both clinical and epidemiological endpoints) and focused on the fact that the estimates of risk for various alternative standards were very similar. The panel concluded that there was no "bright line" that distinguished any of the proposed standards as being significantly more protective of public health. That CASAC Panel focused on the percent responding and the overlap between the estimates given the uncertainty in the estimates rather than on the absolute number of incidences. In contrast, the current SP and CASAC Panel focus on the absolute number of incidences.

---

[6] J. Langstaff, Draft Technical Memorandum, Analysis of Uncertainty in Ozone Population Exposure Modeling, July 24, 2006 at page 42.
[7] J. Langstaff, Technical Memorandum, Analysis of Uncertainty in Ozone Population Exposure Modeling, January 2007 at page 52.
[8] See Brochu et al. Hum. Ecol. Risk. Asses., **12**, pages 675-701 and 731-761 (2006).

There is one other major difference between the two panel's recommendations. The current panel did not consider the extremes of background and so incorrectly assumed that uncontrollable background was substantially lower than thought in the previous review. This led CASAC to assume that a standard as low as 0.060 ppm is attainable and that, therefore, exposures of concern could be essentially eliminated

Epidemiological Studies

EPA and CASAC continue to overestimate the magnitude and consistency and underestimate the uncertainty in the results of acute epidemiologic studies, especially for school absences, emergency room visits for asthma, hospital admissions, and mortality.

The first and main issue in interpreting the epidemiologic associations is whether they are real health effects or not, contrary to the acute respiratory effects in clinical studies where we know the effects are real and the issues are the prevalence and public health significance. Although there clearly are positive associations with ozone in the literature for many health endpoints, there are also positive associations for the same health endpoints for many other pollutants. In addition, in systematic studies the overall pattern of associations is similar for each of the pollutants and the overall range of associations is not biologically plausible.
.
While there are now studies of many relevant endpoints, the choices concerning how to evaluate, display, and interpret this large database are critical. The SP consistently made such choices to make the argument for these associations being real health effects. In doing so, it omits or downplays important findings in the literature and patterns in the data. While there are positive associations for various health endpoints, when all the relevant information is considered, there is much less consistency than claimed in the SP. Plotting selected studies with several or even many positive associations may give the appearance of consistency but it is not a rigorous test.

Two examples are noted here; additional discussion of the epidemiological database is included in Appendix B. First, for premature mortality, in multi-city studies there is an implausibly wide range of associations for mortality from strongly positive in some cities to no effect in some cities to strongly negative (implying protective effects) in some cities. The wide range for ozone is shown in the NMAPPS data-set in Figure 7-17 of the CD.

Second, even though CASAC asked staff to use the data on asthma and other respiratory emergency room visits to support the position that there are significant adverse effects below the current standard, a number of rigorous evaluations of the literature come to different conclusions. Anderson, et al. 1998[9] reported that several pollutants had positive associations with asthma admissions in London, but that there was a lack of consistency across age groups and seasons. Anderson et al. also evaluated 15 other studies of air

[9] Anderson et al. Thorax, **53**, pages 842-848 (1998).

pollution and daily asthma emergency visits and admissions in the literature and concluded "Taken overall, it is apparent that the evidence is not coherent as to whether there is an effect of pollution or the responsible pollutant." They list a number of possible reasons for the lack of consistency, including false negatives due to lack of statistical power and false positives due to chance, multiple significance testing, post hoc hypothesis testing, and publication bias. They also note differences in pollution level and mix between cities, the presence of highly correlated pollutants, and that pollutants acting as surrogates for unmeasured pollutants or ambient aeroallergens may be involved. They conclude that, while there is evidence that all of the pollutants may have an effect on asthma, there is a lack of consistency in the specific pollutant responsible.

The 2004 ozone standard review in California indicated that if hospital admissions are affected by ozone, then it is likely that emergency department (ED) visits would also be affected, and to a greater degree. However, the ED results appeared less consistent, and were summarized as being inconclusive. The California review[10] acknowledged that there is no clear pattern of association with specific respiratory disease outcomes. In the CD, for asthma–related emergency room visits, the range in ozone associations is from –20 % to + 80 % per standard ozone increment as shown in Figure 7-8 . For total respiratory hospital admissions, the range is from – 15 % to + 15 % per standard ozone increment as shown in Figure 7-9. With the wide range of positive and negative associations and the many inconsistencies noted by others, it is hard to accept this data as strong evidence of ozone effects.

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 as causal.

The SP should fully acknowledge that the body of studies since 1996 demonstrates major uncertainties and issues related to model selection and publication bias. There is an implausibly wide range of results from positive to negative in systematic analyses. When the full range of studies in the CD is viewed, the range is also very wide but shifted to more positive findings - an indication of publication bias. This is documented in Appendix B. The heterogeneity implied by these ranges is biologically implausible. During the ozone review, the input from the CASAC panelists varied regarding the use of these data; however, there was much more skepticism regarding causality than was the case in the PM review. This is reflected in CASAC's June 5, 2006 letter to the Administrator that raises serious issues regarding the use of time series studies.[11]

While staff indicates that there is strong evidence supporting the judgement that the relationships are causal for respiratory symptoms in asthmatic children and respiratory hospital admissions, staff acknowledges that there is greater uncertainty regarding

[10] California Air Resources Board Staff Report, Review of the California Ambient Air Quality Standard for Ozone, March 11, 2005,
[11] R. Henderson, CASAC Teleconference Meeting to Provide Additional Advice to the Agency Concerning Chapter 8 of the Final Ozone Air Quality Criteria Document, EPA-CASSAC-06-007, June 5, 2006

mortality, describing the overall body of evidence as highly suggestive.[12] However, as noted in earlier AIR comments, the latest multi-city study of respiratory effects in asthmatic children in eight cities, including Boston, found associations with other air pollutants but not with ozone. In addition, the Mortimer study implicates air pollution in general, not ozone in particular. Therefore, the data for asthmatic effects in children is best described as mixed and inconsistent rather than clear and strong evidence of an ozone effect. The results for respiratory hospital admissions, similarly, are more inconsistent than staff acknowledges.

EPA staff and CASAC also overstate the biologic plausibility of the epidemiologic associations for serious health outcomes

The controlled human studies show 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. As noted above, the evidence for respiratory emergency room visit is inconsistent and inconclusive and the evidence for symptomatic effects in asthmatic children is mixed and inconsistent.

A major area of uncertainty regarding the risk assessment is a set of issues related to the empirically-estimated concentration-response relationships.[13] These include uncertainty related to model selection, shape of the response, and confounding. As documented in Appendix B, the uncertainty in the "true" relation is much larger than the statistical uncertainty in any given association. While staff concluded that a linear response down to background should be applied in the risk assessment, there are many reasons why this should not be the baseline estimate. CASAC rejected this approach for PM even though the PM database was viewed as more consistent.

The CD documents that the day-to-day personal exposures of the population are very low even in areas that substantially exceeded the 1-hour and 8-hour ozone standards in the past. This fact provides a challenge to explain how low ozone concentrations, within the range of background, can be causing the premature mortality estimated in the risk assessment. For example, Figure 5-15 of the SP shows that the bulk of the estimated risk comes from low ambient exposures and the personal exposures of the frail population are much below the ambient levels.

If one takes the mortality 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. The wide pattern is shown in Figure 7-17 of the CD, but downplayed as just heterogeneity due to differences in the cities.

---

[12] Staff Paper, *supra* note 1, at page 5-93.
[13] Id. at page 5-42.

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.

Precision of the Standard

In recommending the ozone standard be specified to the third decimal place, CASAC overstates the precision and accuracy of current ozone measurements.

CASAC recommended that the ozone standard be specified to the third decimal place in ppm because current measurement technology "allows accurate measurements of ozone concentrations with a precision of parts per billion." Although California recently specified an 8-hour ozone standard to the third decimal place in ppm, the Air Resources Board staff acknowledges that the third decimal place is an uncertain digit.[14] At the 0.070 ppm level, Air Resources Board staff indicated that ozone measurements are uncertain in the third decimal place to the degree of plus or minus 0.003 ppm. Another way of expressing this uncertainty is that 95 % of the true values will be within two standard deviations, or plus or minus 0.006 ppm. The EPA estimate of precision noted in the SP of plus or minus several percent is consistent as is the precision reported for the CASTNET sites where 98 % of the audits were within plus or minus 10 %. Given this level of precision, the current round-off procedure is appropriate.

The SP refers to an analysis by Cox and Camalier that purports to show that the impact of measurement error on the design value is small -- the order of 1 ppb for precision and 1.3 ppb for randomly occurring instrument bias. However, there is no reason to assume that the bias at a given instrument is random. Therefore, the staff analysis is not sufficient to support a change to the round-off procedure.

Section II Comments Regarding the Secondary Standard

Background or Uncontrollable Ozone

The extremes of policy relevant background concentrations must be considered in the selection of the level for the secondary standard.

The same analyses, arguments, and data that apply to the primary standard also apply to the secondary standard. As discussed above and in Appendix A, the global model EPA relied on is not capable of resolving the phenomena that contribute to attainment or non-attainment of secondary ozone standards throughout the country. In fact, even the more refined, higher-resolution CMAQ model, which EPA evaluated when developing the "as-is" W126 exposure surface in the West, performed extremely poorly in predicting the observed SUM06 and W126 in 2001. This means that the sources and processes that

[14] California Air Resources Board Staff Report, Review of the California Ambient Air Quality Standard for Ozone, Volume II, March 11, 2005, at page 6-3.

contribute to the substantial levels of both SUM06 and W126 in rural and remote areas in the western U. S. are not well-captured in state-of-the-art chemical transport models.

Since CMAQ was developed to focus on the impact of U. S. man-made pollution sources, the poor prediction performance suggests other sources and processes are responsible for a substantial portion of the elevated SUM06 and W126 levels observed in rural and remote western areas.

The spatial distribution of current W126 levels indicates both traditional anthropogenic as well as biogenic and agricultural emissions and other background sources contribute to elevated W126

The map of "as-is" W126 levels included as Figure 7-6 of the SP shows W126 levels exceeding 7 ppm-hr throughout many areas of the country. Although some areas of elevated W126 occur in and downwind of major man-made source areas, other elevated W126 areas occur in less populous areas including: Utah, New Mexico, Wyoming, Kansas, and Oklahoma, states where population and density of man-made emissions are only a small fraction of that in the eastern U. S.

The overall pattern of W126 suggests both a contribution from U.S. man-made emissions and a background component that could involve biogenic emissions, agricultural emissions, long-range transport from non-U.S. sources, and a stratospheric component. In particular, W126 levels greater than 7 ppm-hr exist in many locations throughout the country, including many rural and remote Western sites well-removed from anthropogenic emission source areas.

Reliance on an inadequate model to estimate background contributed to the CASAC recommendation for secondary standards that are within the range of uncontrollable background.

EPA assumes, incorrectly, based on the Fiore et al. modeling analyses, that SUM06 or W126 is not confounded by background ozone. However, substantial levels of both SUM06 and W126 in rural and remote areas are not obviously related to known man-made source areas.

CASAC's recommendations came after reviewing the second draft SP and should be reviewed and modified. The second draft SP provided data on the distribution of SUM06 with insufficient detail to enable an accurate determination of the spatial extent of non-attainment for standards within the range recommended by CASAC. As noted above, both models and observations must be used when determining the extent of uncontrollable ozone. Otherwise, a secondary ozone standard may be implemented that is unattainable in reality.

In Figure 2, longer-term W126 data are plotted for a series of sites across the Western U. S., from the California/Oregon border across Idaho, Wyoming and into South Dakota. All of these sites are substantially removed from major emission sources and have W126 values generally in the range of 0 to 20. Note also that there are wide annual variations from. There are several important implications from this data.

1) Even sites removed from obvious source areas have a substantial number of hours annually at or above 0.06 ppm, often on the order of 100 to 200 hours per year. Importantly, all these sites have 8-hour design values of 0.06 or greater.
2) The W126 metric varies substantially from year to year even in rural and remote areas lacking any obvious influence from anthropogenic sources. Consequently, if a W126 or other cumulative standard is set using only one year's data, areas could bounce in and out of attainment, causing havoc with implementation planning efforts.

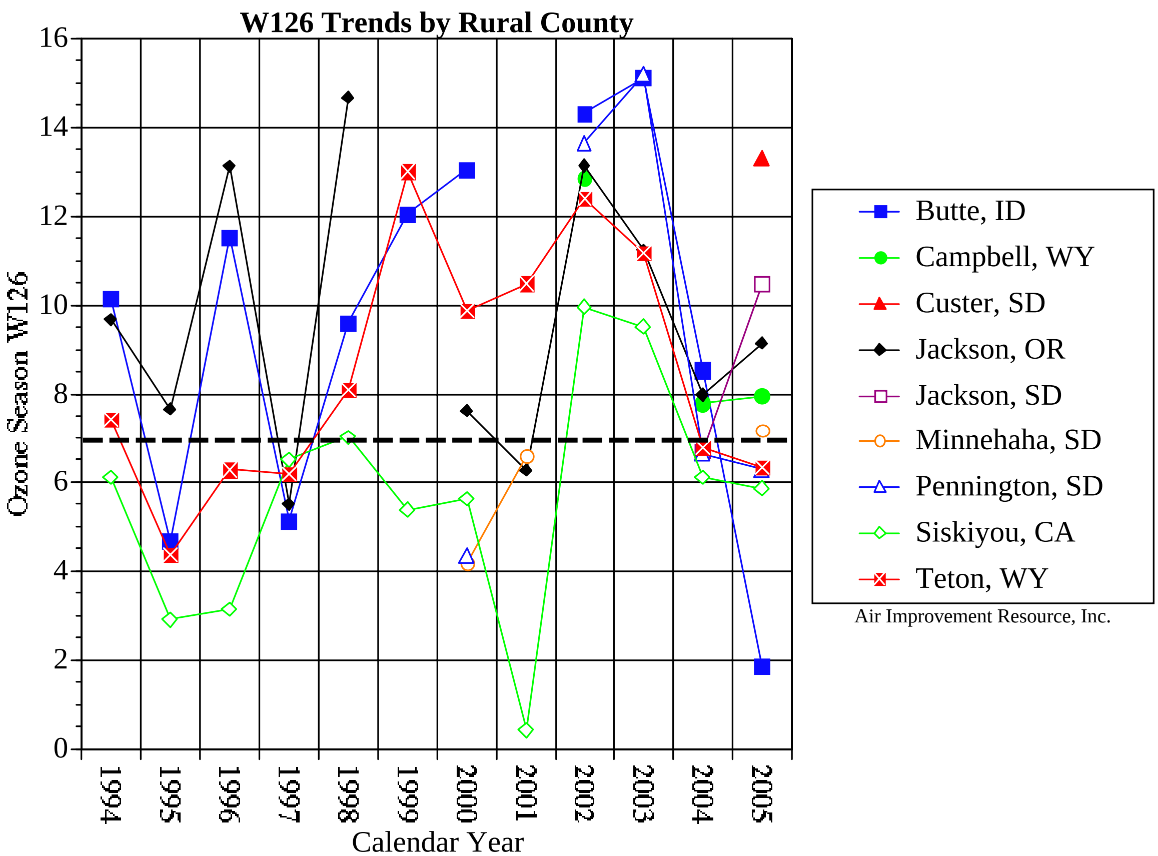


Figure 2

Vegetation Effects

No significant new data on vegetation effects exists since the last review suitable for establishing new secondary standards

The vegetation risk assessment utilizes the crop yield loss and seedling growth data used in the previous review. Due to the lack of new data, the vegetation risk assessment does not materially reduce the major uncertainties that were acknowledged in the 1996/1997 review. The uncertainties in extrapolating from effects in open-top chambers to effects in natural settings and extrapolating from seedling effects to effects on forests are both still major concerns. In addition, there is still a paucity of ozone monitoring in crop-growing regions.

Metrics to Protect Vegetation

The choice of an appropriate ozone metric to protect vegetation is difficult.

Although cumulative ozone metrics offer advantages over the form of the primary standard to protect vegetation, the limitations and simplifications involved with cumulative ozone metrics are still of concern. For example, the simplifying assumptions that plant exposure to ozone equals plant response and that there is a uniform threshold or response function for plant injury are both known to be incorrect. Although more biologically-based flux-based models offer promise, there are insufficient data to apply them at present. In spite of these issues, EPA staff and CASAC recommend a seasonal cumulative secondary ozone standard.

The analysis used to discriminate between cumulative standards and the current 8-hour standard is flawed

A major consideration in the choice of a seasonal standard was the assessment of the extent to which an 8-hour standard would provide protection for vegetation. The assessment involved a comparison of the "as-is" W126 exposure (Figure 7-6 of the SP) with the W126 exposure rolling back 8-hr ozone concentrations to meet the current standard (Figure 7-7 of the SP). This comparison is misleading, because the rollback approach assumes that there will be no reduction in ozone at sites that currently attain the 8-hour standard. In reality, there are many national and regional emission reduction programs that will reduce man-made ozone throughout the country and measures adopted through the SIP process that will also reduce ozone in current attainment as well as non-attainment areas.

The form of the current 8-hour standard, while not designed to protect vegetation, does focus control on the precursors of man-made ozone and consequently will reduce the overall level of man-made ozone throughout the country. EPA's analysis to date does not evaluate the impact of the nation's ozone control program on W126 in a scientifically sound manner. Before a decision regarding the protectiveness of an 8-hour standard is made, a realistic evaluation of the ozone reductions throughout the country must be made using state-of-the-art photochemical models.

The European experience with cumulative standards has been disappointing

While both EPA staff and CASAC recommend a cumulative ozone metric, the European experience with cumulative ozone metrics has been disappointing. In a series of recent papers,[15] European researchers have acknowledged that the AOT40 critical level adopted

---

[15] See the series of papers by Ferretti et al. in Volume 145 of Environmental Pollution, pages 617 –655 (2007) that report the findings from monitoring of forest conditions and ozone at 83 sites in France, Italy, Luxembourg, Spain, and Switzerland over the years 2000-2002, and the series of papers in Volume 146 of

in Europe has proven to be unrealistic and unattainable and have called for revisions to the approach. In addition, it has been reported that visible symptoms at 83 sites in five countries had no relation to the cumulative exposure metric and that crown defoliation of beech was only related to ozone exposure at very high exposure levels, seven times the European critical level for forests.

In fact, in the European approach, the cumulative seasonal metric was chosen as a first step that evaluated the potential risk to vegetation, while recognizing that it does not take into account environmental factors that modify ozone uptake, plant defense mechanisms, or differences in ozone tolerance between species. The Europeans are now working on their second level approach which will be flux-based.

Vegetation Risk Assessment

The estimates of crop loss are substantially lower than those of the previous review

The SP acknowledges that the estimates of crop loss are substantially lower than in the previous review. As a result, the economic benefits associated with attaining various alternative standards have been substantially lowered from previous ozone yield loss studies. It is now widely accepted, that ozone concentrations at the "standard measurement height" are not the same as the ozone exposures at plant height. *Cite*. When this factor was corrected, the SUM06 and W126 exposures were approximately cut in half.

Studies of foliar injury do not support a cumulative metric

The map of locations where foliar injury was present and absent in 2001 (Figure 7-19) shows a patchwork of responses that, in many cases, are not obviously correlated to major areas of man-made precursor emission. Although foliar injury is observed in areas that meet the current 8-hour standard, there are also high and low SUM06 areas that experience foliar injury and areas that do not. Similarly, the European experience indicates that there was little or no relation between a cumulative metric, AOT40, and foliar injury at 83 sites in 5 countries.

At a site in the Northeast corridor of the U. S. where ozone levels are elevated, Davis and Orendovici [16] report that there was no readily apparent relationship between SUM06 ozone levels and incidence of ozone stipple. SUM06, by itself, was not a reliable predictor of ozone foliar injury, but when included with N100 (the number of hours of 0.100 ppm or higher) and drought stress, a cumulative ozone metric such as W126 did have substantial predictive power. Although the SP mentions the possible use of N100 along with a cumulative ozone metric, it took no position on the question. Clearly, this is an important issue that should be evaluated fully by EPA.

Environmental Pollution (2007) pages 577- 770 that contains peer-reviewed contributions from the November 2005 Critical Levels of Ozone Workshop held in Obergurgl, Austria.

[16] Davis and Orendovici, Incidence of ozone symptoms on vegetation within a National Wildlife Refuge in New Jersey, *Environmental Pollution*, **143**, pages 555-564, 2006.

CASAC made its recommendations concerning the secondary standard based on the draft text that contained errors overstating the magnitude of vegetation effects

There were major errors in the translation of the results of the vegetation risk assessment into the second draft SP, overstating the magnitude of crop loss from ozone. For example, the table summarizing the results of the crop loss risk assessment in the second draft SP reported the maximum yield loss for various crops (8 crops and 7 fruits and vegetables) but labeled the numbers as the median yield loss. This grossly overstated both the yield loss from current ozone levels as well as the yield gain from attaining any of the alternative secondary standards under consideration. In the final SP this was corrected. However, the SP still only contains the maximum yield loss figures. The actual distribution of yield loss can only be found in the Appendix G of the TSD[17] which indicates that the mean and median yield gains from attaining any of the alternatives under consideration are quite small – the order of 1 % or less for even the most stringent standards. Unfortunately, CASAC made its recommendations concerning the secondary standard based on the draft text that contained numerous errors overstating the magnitude of vegetation effects; CASAC must review the final SP before making a final recommendation.

The recommended range for a cumulative standard needs to be adjusted upward

The CASAC suggestions for the range of a SUM06 or W126 standard come from a consensus-building workshop held in 1997. The EPA staff's recommended range also considered the results of the 1997 review and risk assessment. Since neither of these sources considered the difference between the ozone monitored in the network and the ozone exposure at plant height, these ranges should be re-evaluated accordingly.

If ozone monitoring is continued under the current guidelines, the ranges for the seasonal secondary standard must be adjusted upward. Based on the reduction in ozone exposure at plant height of a factor of two compared to that at measurement height, the range should be increased by roughly a factor of two. Neither CASAC nor EPA took this into account in making their recommendations concerning the level of the secondary standard.

Implementing a W126 standard

Since EPA staff recommends a seasonal secondary ozone standard, the question of whether monitoring and modeling tools are available to implement such a standard needs to be addressed by EPA and CASAC.

The placement of current ozone monitors is not appropriate for evaluating vegetation exposure in terms of either spatial coverage or monitor height. Therefore, a new monitoring network must be designed and implemented before designations of attainment or nonattainment occur. Based on work to date, a model suitable for developing control

[17] Lehrer et al., *supra* note 4, at Appendix G.

strategies for a cumulative seasonal standard, particularly in the western U. S., is not available. In addition, major improvements would be needed in grided emission inventories throughout the U. S. particularly for agricultural emissions. Therefore, the tools needed to implement a seasonal ozone standard are not in place.

When the timeframe required to implement a new monitoring network, to collect data for several years, and to develop and approve an implementation plan is considered, the next ozone review will have been completed before implementation plans will be in place. Ozone reviews are required every five years and EPA recently changed the review process dramatically to meet the statutory requirement. As an example of the timeframe needed to implement a standard when a new monitoring network is required, the PM2.5 standards were set in 1997 and the state implementation plans are due to EPA in 2007.

All this assumes that tools are available when needed, sources responsible for the nonattainment can be identified, and the standard is attainable in the sense that reduction or elimination of the identified sources is at least theoretically possible. To the extent that any of these steps are not possible, the setting of a seasonal secondary standard will not be fruitful and will not aid materially in the protection of vegetation.

Section III Comments on the Role of CASAC

In addition to advising the Administrator on the health and welfare aspects of the standard, CASAC is charged with the responsibility for identifying background ambient levels as well as oversight of the SIP process for both primary and secondary standards. These responsibilities have not been fulfilled.

One of the specific charges for CASAC under the Clean Air Act (Section 109 (d) (2) (c) (iii)) is to advise the Administrator on the relative contribution to air pollution concentrations of natural as well as anthropogenic activity. This is an extremely important issue in the case of ozone, particularly since public comments have raised the issue that the ranges recommended by EPA and CASAC overlap with the uncontrollable ozone background. A complete reading of the CD indicates that the SP's reliance on one particular global model is not a scientifically sound way to estimate the contribution of uncontrollable ozone to an extreme value primary or a cumulative seasonal secondary standard.

Another specific charge to CASAC (Section 109 (d) (2) (c) (iv)) is to advise the Administrator of any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of national ambient air quality standards. Thus, CASAC is charged with oversight of the SIP process for both primary and secondary standards.

Since EPA and CASAC are recommending a separate secondary standard that will require a major effort to implement, it is incumbent on CASAC to consider the ramifications of its recommendations in the implementation phase and evaluate the availability and limitations of tools to implement a separate secondary standard.

In addition, CASAC should evaluate whether the full ranges recommended for both the primary and secondary standards are attainable.

The current Ozone Panel is not equipped to carry out these charges

The membership of the current panel is weighted heavily towards expertise in health and welfare effects. Additional expertise in other areas would be needed to fully carry out the two charges noted above. Before the nation embarks on a major tightening of the primary ozone standard and the promulgation of a separate secondary standard, the question of whether the ranges recommended by EPA and CASAC overlap the background of ozone uncontrollable by reduction of U. S. emissions needs to be fully addressed and answered.

Appendix A Detailed Comments on Background Ozone

Issues related to the use of global models to estimate peak ozone background

Many scientific issues with the use of global chemical transport models are discussed in section AX2.5 of the CD. These include the difficulty of any model capturing tropospheric 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. 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. 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. 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 influence peak ozone levels and 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.

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 from long-range transport from outside the U. S. The CD documents the limitations of the Fiore et al. 2003 modeling study as well as other global models at simulating these mechanisms. In order to evaluate the range and variability in PRB, the Agency and CASAC need to consider both observations and models. Since the ozone standards have historically been defined as yearly extremes of daily maxima, the SP should have evaluated 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:[18]

> "…a reasonable estimate of the background O_3 concentrations near sea level in the U. S. for a 1-hour daily maximum during the summer is usually in the range of

[18] 1996 OAQPS Staff Paper at page 20.

> 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."

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 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. For example, Winner and Cass modeling the Los Angeles Basin over an entire year and evaluated VOCand NOx controls from 0 to 95 %, They concluded that the current 8-hour standard may be unattainable in the basin even with complete elimination of man-made emissions and that the days that are most difficult to control are not necessarily the days that produced the peak concentrations in the historical base year.[19]

It is particularly important that the qualifications and limitations concerning background in the CD be considered as the Agency goes forward to propose any revisions to the ozone standards. For example, section AX3-9 of the CD, which discusses the methodology for evaluating background, specifically notes that the late winter and spring increase in ozone at remote sites is caused by a combination of intercontinental transport and stratospheric input, that Eulerian models such as GEOS-CHEM cannot resolve or characterize the impact of individual stratospheric intrusion events, that global scale models cannot simulate the effects of such intrusions on surface ozone concentrations, that the coarse resolution precludes simulation of fine-scale structures or localized extreme events, that GEOS-CHEM cannot capture the structure and enhancements associated with stratospheric intrusions, and that GEOS-CHEM underestimates sub-grid-scale local peaks.   The same section notes that there are a variety of processes that can cause elevated ozone in rural and remote areas, including intercontinental transport, stratospheric intrusions, wildfires, and regional transport. To rigorously discriminate among these sources, ancillary measurements and/or analyses are needed. Unfortunately, EPA has not focused its resources on measurements of ozone and ancillary variables in agricultural or other rural and remote areas, so the research needed to identify the various contributions to elevated ozone from controllable or uncontrollable U. S. sources has not been carried out.

As we noted in 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

---

[19] D. Winner and G. Cass, "Effect of Emissions Control on the Long-Term Frequency Distribution of Regional Ozone Concentrations," Environ. Sci. Technol., **34**, pages 2612-2617 (2000).

Chapter 2 of the CD. We also pointed out that Fusco and Logan 2003[20] 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 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 of interpreting any global model 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 NOx 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 (2^o x 2.5^o) 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 represent 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 CD indicated that "local maxima and minima are difficult to reproduce with global models since processes are averaged over an entire model grid cell."[21] Liang and Jacobsen[22] have shown that integrated ozone production may be over- or under-predicted when coarsely resolved models blend air masses of different emissions character. Over-predictions by as much as 60 % were found for coarse-model grid cells applied 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 from rural and urbanized areas together in large grids, it clearly does not simulate the concentration fields of precursors correctly on the scale that influences urban plumes and daily ozone peaks. So even if the model predicts the correct ozone, it is doing it for the wrong reason. Therefore, the model's predictions for altered emission profiles are also suspect.

The CD indicates that the simulation of stratospheric intrusions is notoriously difficult in global models (AX3-144). This arises because stratospheric-tropospheric exchange is an intermittent process, a series of strong exchange events separated by more quiescent

---

[20] 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).
[21] CD, *supra* note 3, at page AX2-119.
[22] J. Liang and M. Jacobson, "Effects of subgrid segregation on ozone production efficiency in a chemical model," Atmospheric Environment, **34**, pages 2975-2982 (2000).

periods.[23] 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.

Since the current ozone standard has an extreme value statistical form, 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 CD indicates many sources of uncertainty and variability in addressing that task. For all these reasons, the Agency must not rely solely on this one modeling study to define the uncontrollable background.

The CD notes the need for additional work to evaluate PRB with an ensemble of models comparing the model results with observations.[24] The CD notes the need to include models with greater resolution to understand variability on shorter time scales and variability due to processes that are not captured in the global models. Until this is done, a range of background estimates from models and observations should be used by EPA. Because of the acknowledged limitations of global models, observations should be the primary source for estimates of extreme values of background at specific locations.

In addition to the concern with the use of an inappropriate tool to estimate peak background, we continue to have major concerns with the way EPA defines background. The Agency continues to define 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 earlier AIR comments, 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.

Issues related to the use of observations to estimate background

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

[23] A. Stohl et al., "Stratospheric-tropospheric exchange: A review, and what we have learned from STACCATO," J. Geophys. Res., **108 (D12)**, page 8516 (2003).
[24] CD, *supra* note 3, at page 3-55.

ozone. However, it goes on to state that PRB estimates must be based on modeling. This is not true. While the 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 related to U. S. man-made emissions. As documented in Annexes 2 and 3 of the CD, there are still major uncertainties and limitations involved in using atmospheric models. These concerns as they relate to the use of global models to estimate background are documented above. The question is not whether to use observations or models, both observations and models should be used to estimate uncontrollable background.

In addition, uncontrollable background needs to be evaluated for each ozone metric being considered for the primary and secondary standards. Since extreme value forms of the primary ozone standard are recommended in the SP, the short-term extremes of uncontrollable background must be evaluated by the Agency and CASAC. Since cumulative seasonal forms that focus on elevated ozone concentrations are recommended for the secondary standard, the extremes of uncontrollable background on those metrics must be evaluated.

As acknowledged in the CD and in the Fiore et al. papers the GEOS-CHEM model is not suitable for evaluating short-term extremes. Therefore, observations must be one of the tools used to evaluate extremes of 8-hour ozone.

For the cumulative seasonal metrics, an important issue is whether any model can accurately predict SUM06 or W126. When the CMAQ model was used to predict SUM06 and W126 in the Western U. S. for the year 2001, it failed miserably. For SUM06, the mean normalized bias and mean normalized error were each over 400 % in the Western U. S. excluding California and over 200 % in California. The performance for W126 was somewhat better but the mean normalized bias and mean normalized error were each over 100 % in the Western U. S. excluding California as well as in California.[25] The implication of this failure is that there is a substantial amount of ozone in the 0.06 to 0.07 ppm concentration range in the Western U. S. that is not predicted by EPA's state-of-the-art regional photochemical model. Therefore, available models cannot be used to evaluate uncontrollable ozone for the seasonal cumulative metrics in the western U. S. and observations must be used.

As noted above, a variety of processes can cause elevated ozone in rural and remote areas, including intercontinental transport, stratospheric intrusions, wildfires, and regional transport. There has been considerable work evaluating the role of intercontinental transport and stratospheric intrusions in recent years utilizing satellites, instrumented aircraft, and lidar to collect relevant data. As noted above, the simulation of stratospheric intrusions is notoriously difficult in global models because stratospheric-tropospheric exchange is an intermittent process - a series of strong exchange events separated by more quiescent periods. These events insert stratospheric air high in ozone into layers or filaments in the troposphere. Although these intrusions penetrate to ground level intact

[25] See Tables A-2 through A-6 of Lehrer et al., *supra* note 4.

only rarely,[26] they intermittently insert high concentrations of ozone into the troposphere much more often.

Elbern et al. report that globally there are about 11,000 tropopause fold days per year.[27] There are spatial and seasonal differences in the number and strength of folds that insert stratospheric air high in ozone into the troposphere. Numerous studies have evaluated the fate of these plumes or layers and found many different patterns in specific instances. Newell et al. report that up to one-fifth of the lowest 12 km of the atmosphere is occupied by quasi-horizontal fine laminar structures of different trace gas composition.[28] Some of these layers or filaments are from continental pollution, some are from stratospheric air inserted into the troposphere, and some are from convection from the boundary layer. There is evidence that some of these layers persist for more than a week and thus can be transported long distances.[29] The layers or filaments of stratospheric origin and intercontinental transport eventually mix with surrounding air in ways which are not yet fully understood, influencing ozone levels, photochemical activity and lifetimes of various species. Wimmers and Moody point out that the fine scale structures observed in the troposphere rapidly collapse to the model grid scale in simulations, and they point out that this largely explains the continuing disparity between models and observations of persistent layers of stratospherically enhanced air in the troposphere.[30] Mixing not only destroys the fine structure but also changes the composition and photochemical activity of the resulting mixture. In rough terrain, there are additional complications and issues. Studies of the transport of urban plumes from California reveal a range of processes involved including terrain following flow at night that inserts plumes and air parcels in the middle and upper troposphere.[31] Thus, global models cannot be expected to capture the features and processes that lead to specific local peaks of uncontrollable background observed at ground level.

As noted above, EPA has not focused its resources on measurements of ozone and ancillary variables in agricultural or other rural and remote areas, so the research needed to identify the various contributions to elevated ozone from controllable or uncontrollable U. S. sources has not been carried out. The SP makes the argument that the conditions conducive to high man-made and high uncontrollable ozone generally do not coincide but that is not necessarily true and it is not the correct issue for an extreme value standard. There is now evidence from airborne lidar that high-ozone air from a tropopause fold can cap a pollution plume and influence boundary layer processes in ways that current chemical transport models cannot reproduce.[32] In addition, Wolff et al. have reported that ^{7}Be, a tracer of stratospheric air, is higher on the backside of high pressure systems where man-made ozone production is also higher.[33]

---

[26] Although direct transport to the surface is rare, there are numerous reports in the literature of ozone levels above 0.08 and up to 0.25 ppm observed at ground-level sites from stratospheric intrusions.
[27] Elbern et al. Theoretical and Applied Climatology, **59**, pages 181-200 (1998).
[28] Newell et al., Nature, **398**, 316-319 (1999).
[29] Bithell et al., Atmos. Environ., **34**, 2563-2570 (2000).
[30] Wimmers and Moody, J. Geophys. Res., **109**, D19306, 4145-4156 (2004).
[31] Kim and Stockwell, Atmos. Environ., **41**, pages 2319-2334 (2007).
[32] Cho et al., Geophys Res. Lett., **28**, 3243-3246 (2001).
[33] Wolff et al., Geophys. Res. Lett., **6**, page 637-639 (1979).

Background ozone estimated from observations

In previous ozone reviews, both the means and extremes of background were evaluated. For example, the Agency's Staff Paper in the previous review concluded:[34]

> "…a reasonable estimate of the background O_3 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."

Although the CD presents and discusses the extremes of ozone at background sites, the current SP ignores the findings of several researchers and previous conclusions of the U. S. EPA regarding maximum background levels. Recently, Vingarzan[35] reported the range of annual maxima at background stations in the U. S. and Canada. The annual maxima (1-hour) at Western U. S. and Canadian background sites approached or exceeded 0.08 ppm.

The level of background ozone first became a policy issue in 1971 when EPA set the first National Ambient Air Quality Standard for Photochemical Oxidants at a concentration of 0.08 ppm for 1-hour, not to be exceeded more than once per year. In 1978, Singh et al.[36] reported an analysis of long-term ozone data from remote sites that indicated summertime average daily 1-hour maxima in the 40 to 50 ppb range but maximum 1-hour concentrations that can approach or exceed 80 ppb in the spring. Singh et al. concluded that achievement of a yearly 1-hour ozone standard of 80 ppb may be impossible.

When EPA revised the federal 1-hour ozone standard in 1979 to 0.12 ppm, the Agency acknowledged that:[37]

> "Field measurements at some remote sites, where man-caused ozone is likely to be negligible, have shown low-but not insignificant- rates of exceedances of the 0.08 ppm level originally proposed for the secondary standard."

In 1989, Logan[38] reported the results of an analysis of ozone data from rural locations in the U. S. She reported that ozone concentrations above 80 ppb were common in rural areas of the eastern U. S. in spring and summer (occurring between about 2 and 8 % of the time) but were unusual at remote western sites, occurring less than 0.5 % of the time. She also pointed out that concentrations of NOx in rural areas of the east are frequently high enough to permit significant photochemical formation of ozone during favorable weather conditions, but that NOx concentrations are much lower in remote regions of the

---

[34] 1996 OAQPS Staff Paper at page 20.
[35] R. Vingarzan, "A review of surface ozone background levels and trends," Atmospheric Environment, **38**, 3431-3442, 2004.
[36] H. B. Singh, F. L. Ludwig and W. B. Johnson, "Troposheric ozone concentrations and variabilities in clean remote atmospheres," Atmos. Environ., **12**, page 2185, 1978.
[37] 44 Fed. Reg. 8212, February 8, 1979.
[38] J. A. Logan, "Ozone in Rural Areas of the United States," J. Geophys. Res., **94**, D6, page 8511, 1989.

west. Logan also reported that the median ozone concentrations of 30 to 40 ppb were similar at rural sites across the country even though there is a much greater population and emission density in the eastern U. S. than in the western U. S.

Lefohn and Foley[39] reported in 1991 on an analysis of ozone data from 26 Class I national parks and wilderness areas. For the seven cleanest sites, the yearly maximum 1-hour average concentrations were in the range of 0.06 to 0.075 ppm.

In 1996, Altshuller and Lefohn[40] published an analysis of background ozone in the planetary boundary layer of the U. S. They selected 11 sites for analysis. The criteria they used included using sites receiving the cleanest air masses from the upwind flow off a continent or ocean, and sites isolated from the influence of urban plumes or regional ozone formation from anthropogenic emissions. They reported that the maximum 1-hour concentrations in the western United States in the April through October period ranged from 50 to 98 ppb and the maximum 1-hour concentrations at coastal sites ranged from 44 to 80 ppb.

In 2001, Lefohn et al.[41] reported an analysis that concluded that hourly levels of 0.05 and 0.06 ppm are often exceeded at northern U. S. and Canadian sites specifically selected to minimize anthropogenic influences. Most occurrences were in April and May but sometimes as late as June. They concluded that many of these occurrences were associated with non-U. S. ozone sources, particularly upper tropospheric and stratospheric air. Although the Fiore et al. 2003 modeling study was carried out to challenge the Lefohn et al. 2001 study, the limitations of the global model noted above preclude its use for estimating the extremes of non-U. S. background, including the extremes of the contribution of stratospheric air.

To demonstrate the range of ozone concentrations present in rural and remote western U. S. sites, we have plotted the distribution of hourly ozone by month for several such sites. Figures 3 through 5 demonstrate that there is a wide distribution of ozone with a broad peak centering around the spring months and with 99th percentile and more extreme ozone concentrations of 0.06 ppm and above. This pattern is observed at rural and remote sites in Wyoming, Colorado, North Dakota, Utah, and other western states. At sites such as Yosemite National Park, where there is well-established transport from urban areas, there is both a rise in ozone levels in the spring but also a broad summer peak indicative of transport of anthropogenic ozone.

[39] A. S. Lefohn and J. K. Foley, "Estimated Surface-level Ozone Exposures in Selected Class I Areas of the United States," paper 91-144.2, presented at the 84th Annual Meeting of the Air and Waste Management Association, Vancouver, British Columbia, June 1991.
[40] A. P. Altshuller and A. S. Lefohn, "Background Ozone in the Planetary Boundary Layer Over the United States," J. Air & Waste Manage. Assoc., **46**, page 134, 1996.
[41] A. Lefohn, S. Oltmans, T. Dann, and H. Singh, "Present-day variability of background ozone in the lower troposphere," J. Geophys. Res., **106**, D9, 9945-9958, 2001.

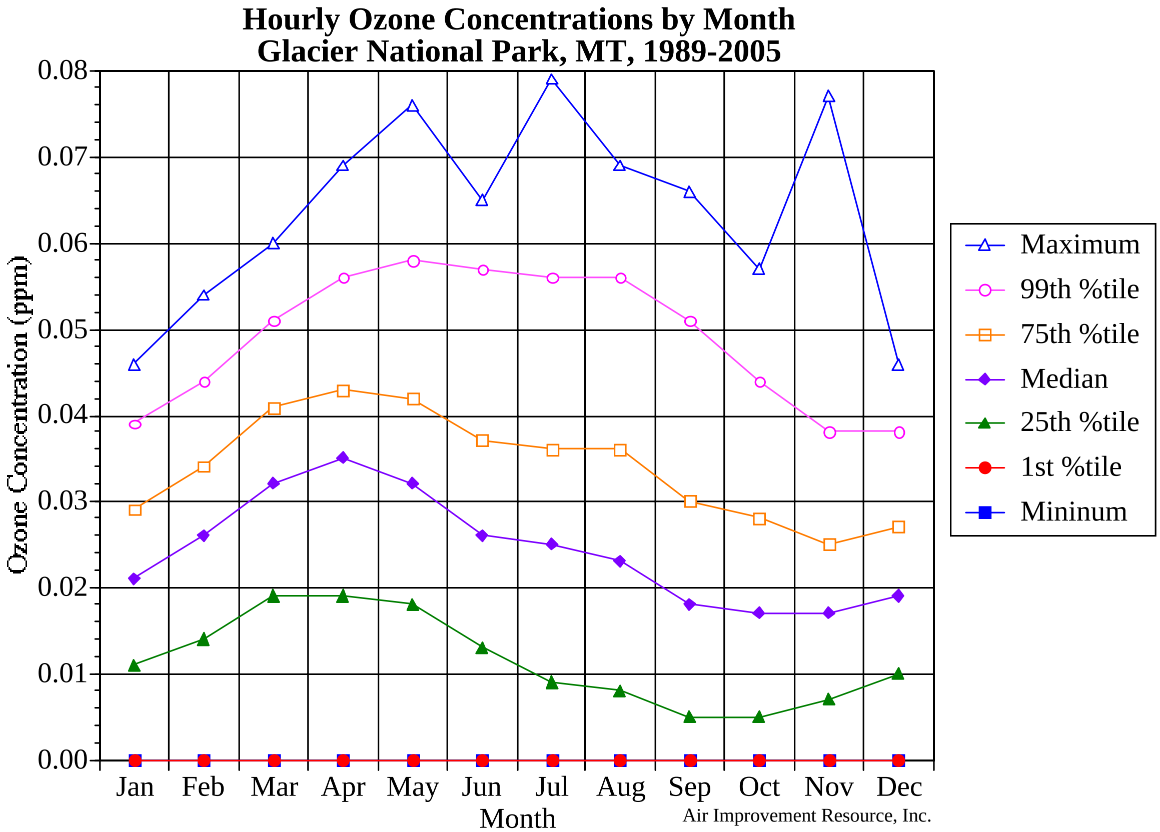

Hourly Ozone Concentrations by Month
Glacier National Park, MT, 1989-2005
Ozone Concentration (ppm)
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
Air Improvement Resource, Inc.
Maximum
99th %tile
75th %tile
Median
25th %tile
1st %tile
Mininum

Figure 3

Figure 4

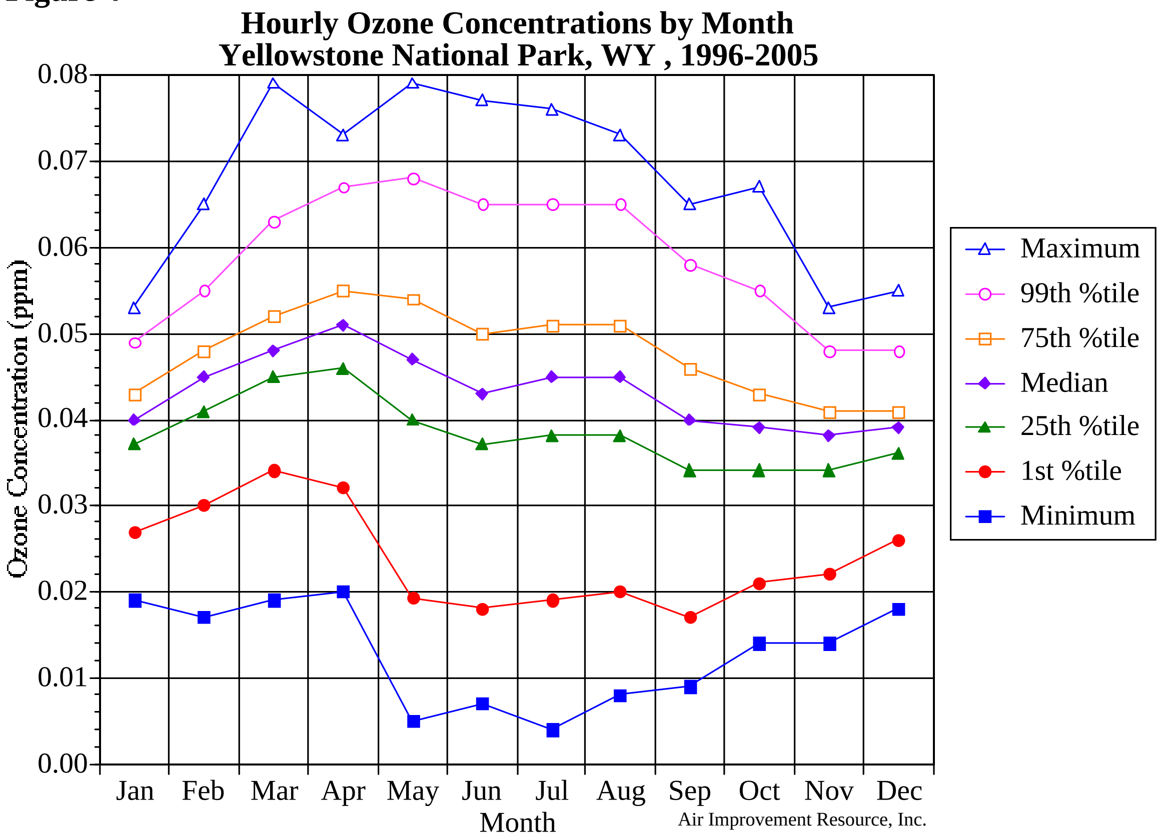
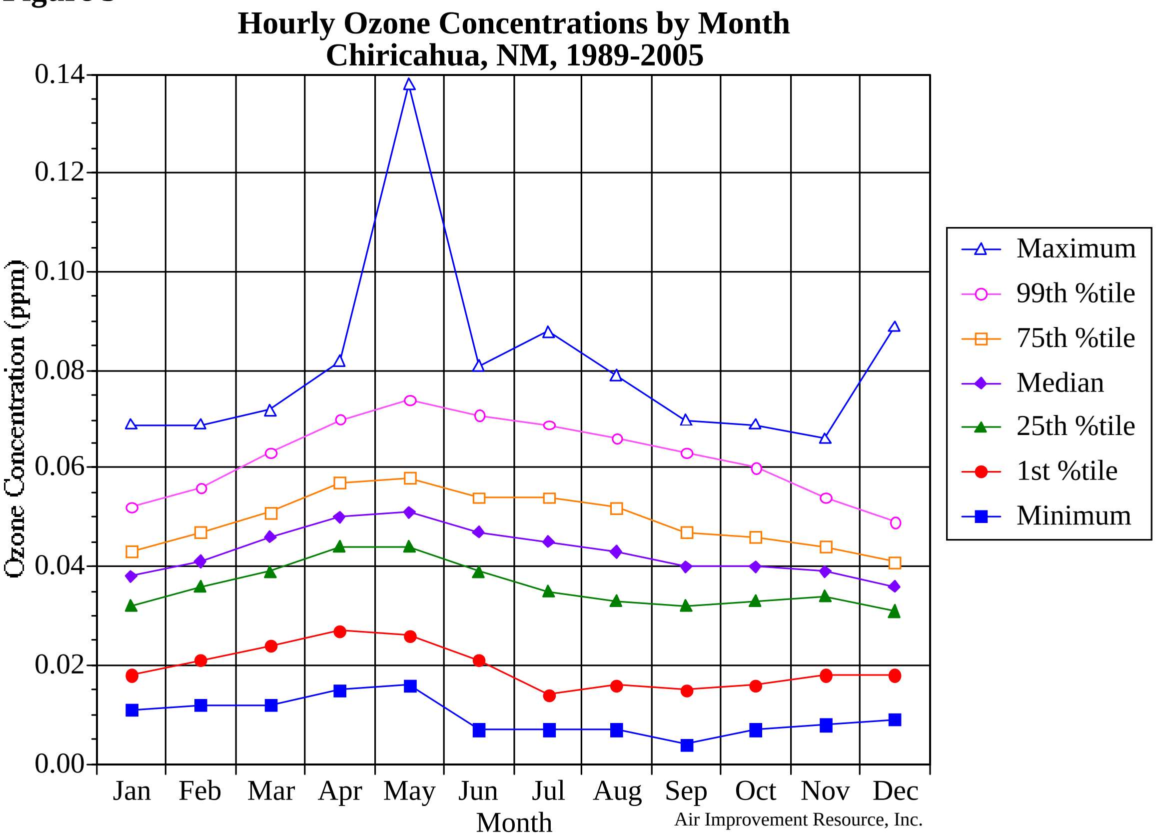

Hourly Ozone Concentrations by Month
Yellowstone National Park, WY , 1996-2005
Ozone Concentration (ppm)
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
Air Improvement Resource, Inc.
Maximum
99th %tile
75th %tile
Median
25th %tile
1st %tile
Minimum

Figure 5



Implications of the extremes of uncontrollable background

With a policy relevant background that varies substantially, there will be times and places where the background approaches or exceeds the levels recommended by CASAC. The SP attempts to limit the discussion of policy relevant background to the meteorological conditions conducive to peak urban ozone formation. While this is currently the limiting case for development of control plans, it would not be under the standards recommended by CASAC.

To illustrate, if the background is 0.040 ppm and the standard is 0.060 ppm, the amount of ozone that can be formed from man-made emissions is only 0.020 ppm. So even with a 0.040 ppm background, a 0.060 ppm standard would allow little room for human activities. On a day when the background is 0.060 ppm, there would be no margin for human activities. While this illustration over-simplifies the complex chemical and meteorological processes involved in ozone formation and transport, it demonstrates that transport of ozone from upwind natural and non-U. S. man-made sources can make some of the standards under consideration unattainable.

The uncontrollable background comes into play in several ways in the setting and achieving of ozone air quality standards. In the SP, mean background is used in the health and risk assessments. Background also plays a role in the designation of attainment or non-attainment, in the development of State Implementation Plans, and in the ultimate achievability of any specific standard. For designations, the appropriate metric is background on the same statistical form as the standard. That is why it is critical that the EPA evaluate the extremes of background for each metric under consideration.

Once a standard is set and designations of attainment/non-attainment are made, states must prepare State Implementation Plans to achieve the standard. Under the federal Clean Air Act, Section 172 (a) (2) (A), the attainment deadline for an area designated nonattainment of a primary standard is "as expeditiously as practicable, but no later than 5 years from the date such area was designated nonattainment." There is also a provision that the Administrator may extend the attainment date for a period of no greater than ten years based on the severity of nonattainment and the feasibility of pollution control measures. These are the deadlines under Subpart 1 of the Clean Air Act that specifies general requirements. There is also a Subpart 2 that provided additional provisions for ozone nonattainment areas with a series of separate deadlines and control requirements based on the degree of nonattainment of the 0.12 ppm 1-hour standard. EPA's framework for implementing the 0.08 ppm 8-hour standard is a mixture of the two approaches with the Subpart 1 approach applying to newly designated areas that were not included as Subpart 2 nonattainment areas.

Congress assumed that air quality standards would focus controls on man-made pollution and would be achievable in reasonable time frames. The penalty for failure to meet by the designated deadline is a bump-up to a higher category with a longer deadline but also more prescriptive control requirements in the control region. The penalty for failure to plan properly is federal sanctions.

CASAC apparently assumed that an 8-hour standard of 0.060 to 0.070 ppm was ultimately achievable from control of U. S. man-made emissions. As documented above, it is not at all clear that this is the case. One of the specific charges for CASAC under the Clean Air Act (Section 109 (d) (2) (c) (iii)) is to advise the Administrator on the relative contribution to air pollution concentrations of natural as well as anthropogenic activity. A complete reading of the CD indicates that the use of one particular global model is not a scientifically sound way to estimate the contribution of uncontrollable ozone to an extreme value standard.

Another specific charge to CASAC (Section 109 (d) (2) (c) (iv)) is to advise the Administrator of any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of national ambient air quality standards. Thus, CASAC is also charged with oversight of the SIP process. Background plays a role in SIP development because of boundary conditions and initial conditions used in photochemical modeling. Over time, photochemical grid modeling has moved from the modeling of multi-day episodes in specific urban areas to modeling ozone concentrations throughout a season over large regions. For the latter situation, an accurate depiction of stratospheric/tropospheric exchange and subsequent transport is necessary. As noted above, the simulation of stratospheric intrusions is notoriously difficult in global models because stratospheric-tropospheric exchange is an intermittent process - a series of strong exchange events separated by more quiescent periods. These events insert stratospheric air high in ozone into layers or filaments in the troposphere. Although these intrusions penetrate to ground level intact only rarely, 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 GEOS-CHEM model is not capable of modeling these phenomena accurately because of its large grid sizes. Because of the complexity of these phenomena, an exceptional events policy would not be able to identify the contribution of stratospheric intrusions to ground-level ozone. The GEOS-CHEM model is also not capable of modeling the extremes of the contribution of long-range transport of ozone from outside the U. S. to specific monitors.

When the current federal 8-hour standardwas set, the administrator considered how close an 0.07 ppm standard would be to background, noting that "As many commenters have noted, based on information in the Criteria Document with regard to ambient concentrations of O_3 from background sources, an 8-hour standard set at a 0.07 ppm level would be closer to peak background levels that infrequently occur in some areas due to nonanthropogenic sources of O_3 precursors, and thus more likely to be inappropriately targeted in some areas on such sources."[42]

If an 8-hour ozone standard were established in the range recommended by CASAC, virtually the entire country would be designated nonattainment. It would be impossible to develop rational implementation plans for some areas and, for other areas, complete elimination of man-made emissions might be required to demonstrate attainment. The extent of non-attainment of an 8-hour standard at the low end of the range recommended by CASAC is shown below in Figure 6. For the most recent data, 2004-2006, a standard at the low end of the range recommended by CASAC would results in 1156 monitoring sites in non-attainment and only 51 in attainment.

In previous AIR comments we had shown that there are many western sites with 4th highest 8-hour averages 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 NOx), 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. Another factor that may lead to lower ozone design values in the northern-most western sites is that the ozone monitoring season is shorter in most of these states.

[42] 62 Fed. Reg. 38868, July 18, 1997.

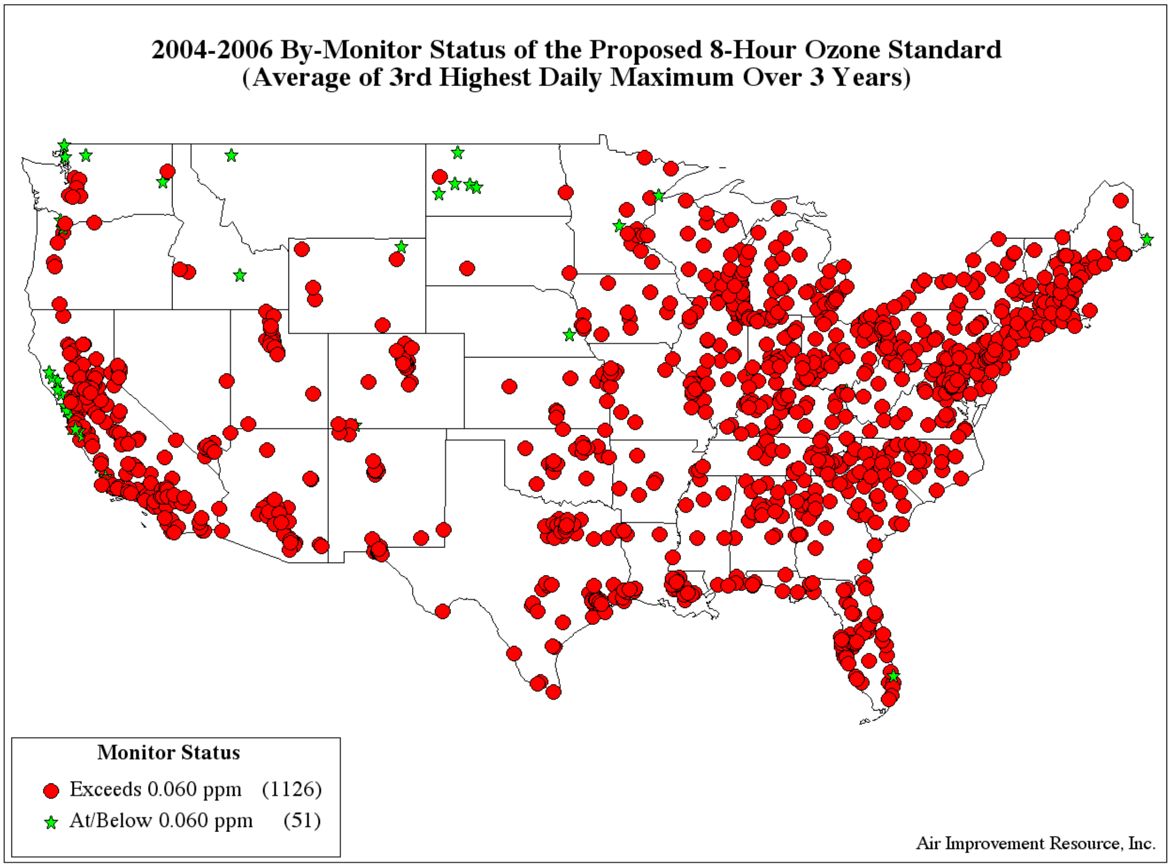


Figure 6

Findings from the interpretation of real world emissions variability

There are three real world aspects of ozone formation in the U. S. that have a bearing on the issue of uncontrollable background. The first is the large variability in population and emissions density that occurs across the country. The second is the long-term trend in emissions that has been occurring for over 30 years, reducing all three ozone precursors - NOx, VOC, and CO. The third is the variation in precursor emissions that occurs between weekdays and weekends. Thus, there is substantial variation in man-made emissions both spatially and temporally. There is also extensive ozone monitoring that can be used to evaluate the impact of these three real-world experiments on the distribution of ozone concentrations across the country. Each of these real-world variations will be discussed in turn.

First, the dramatic variation in population and emissions density between the Eastern and Western U. S. can be used to simulate the impact of reducing man-made emissions by well over 95 percent. A point made on page 3-9 of the CD is that it is hard to find sites in the eastern half of the 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 distribution of ozone concentrations including peak 8-hour values at western monitoring sites can offer one way to estimate uncontrollable background and may provide a practical limit to how low an ozone standard can be before it conflicts with background.

Figure 7 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 7

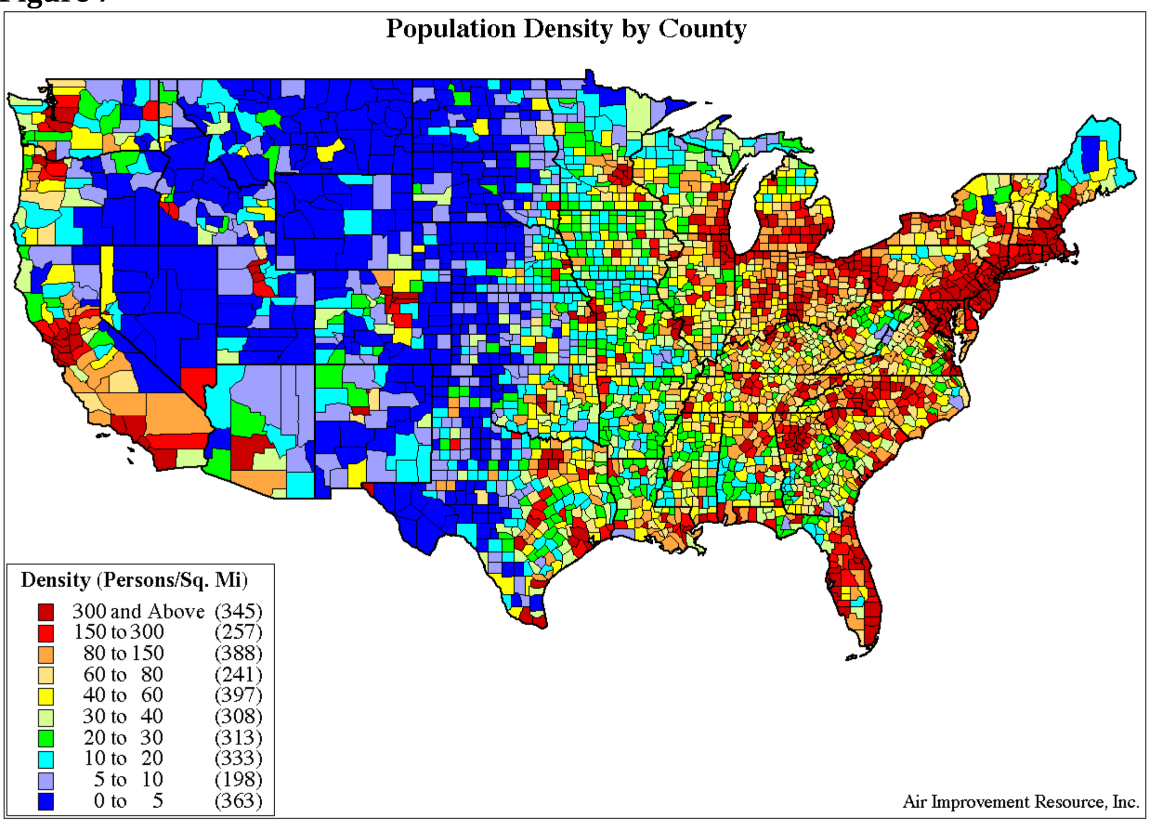


Figure 6 above shows that an 8-hour standard set as low as 0.060 ppm would put virtually the entire country out of attainment leaving essentially no room for ozone from human habitation. Figures 6 actually underestimates the extent of nonattainment of an 0.060 ppm standard because states would need to take year-to-year variability in peak ozone into account and target their control programs below the standard in order to avoid bouncing out of attainment due to meteorological variability.

Figures 8 and 9 show the 8-hour design values for the rural CASTNET sites across the country for 2000-2002 and 2003-2005. In the earlier time period, there is a clear distinction between the rural ozone design values in the eastern and western U. S. as is expected due to the dramatically higher population and emission densities in the two regions of the country. The difference is the order of 0.01 to 0.015 ppm. In the 2003 to 2005 time period, the difference is substantially reduced, presumably because of the substantial emission reductions from the NOx SIP Call. The small difference in peak 8-hour levels in areas of the country with very large differences in man-made emission densities suggests the presence of significant extremes of uncontrollable background in areas with very low emissions.

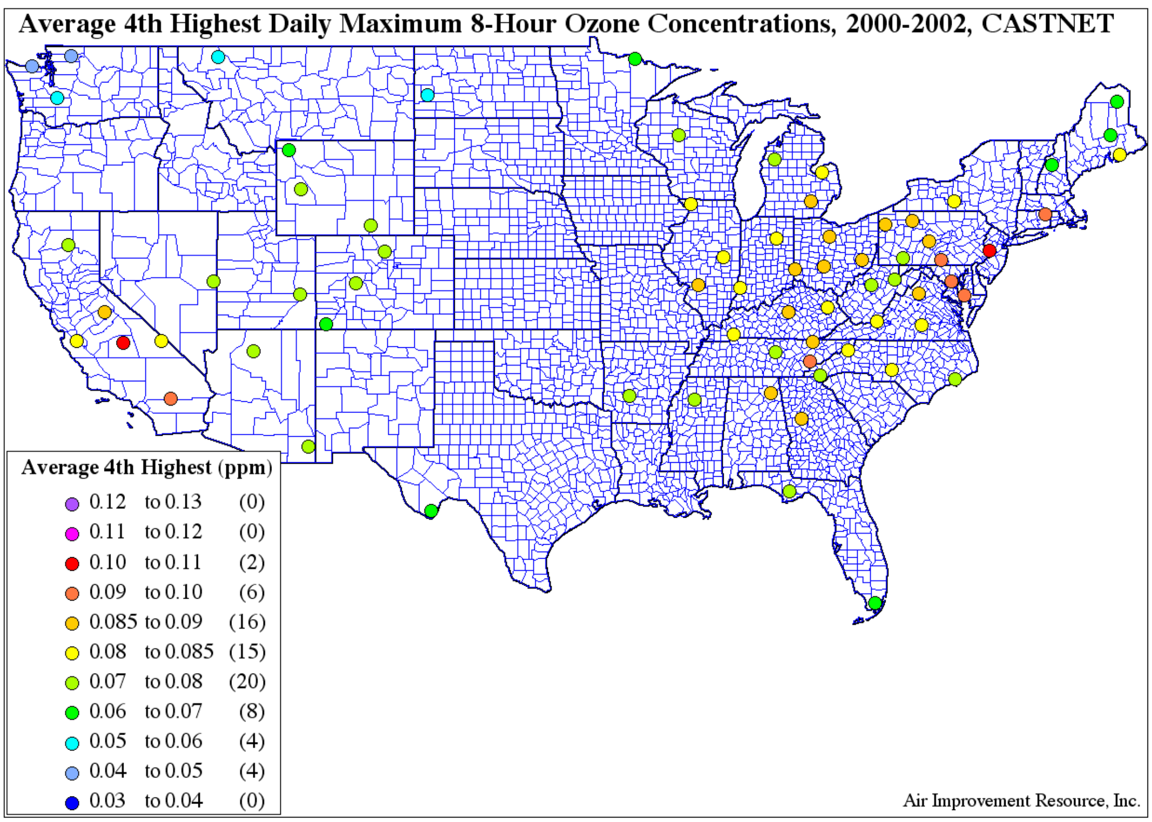


Figure 8

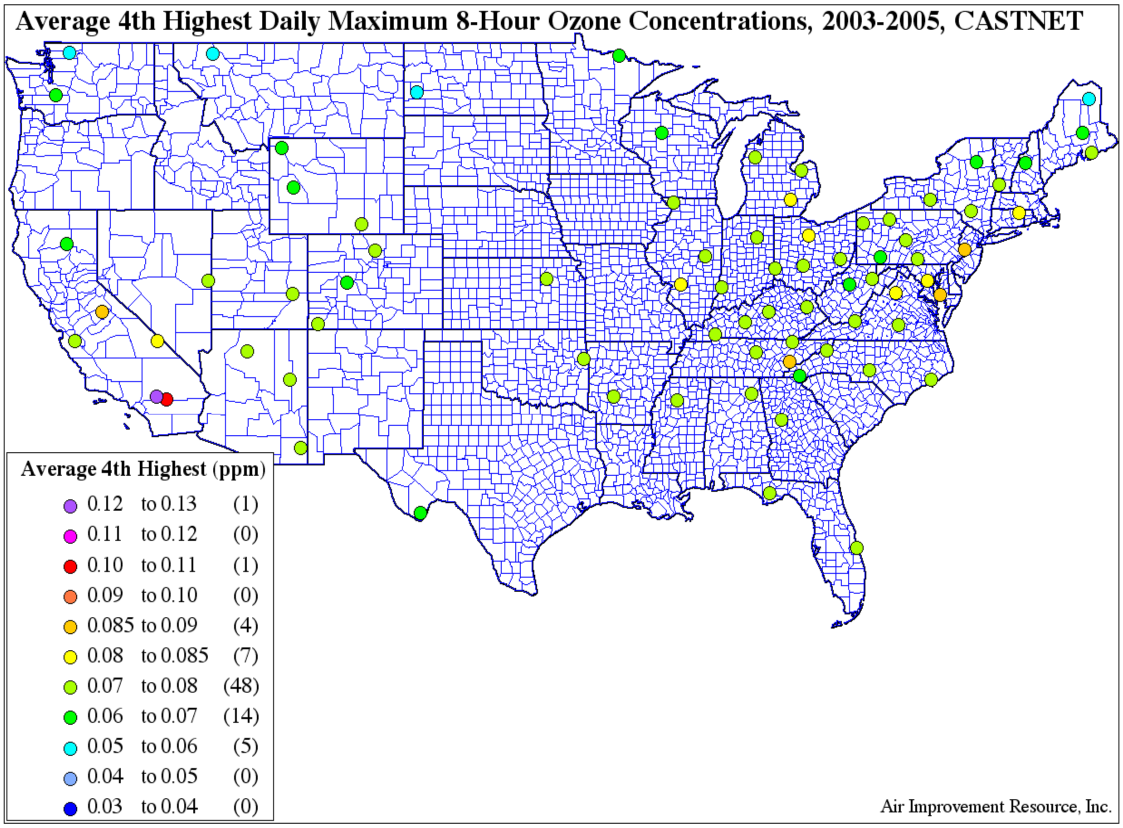


Figure 9

Figures 2-6 and 2-7 of the SP displays the 2002-2004 peak 1-hour ozone and 8-hour levels across the country. When viewed in relation to the population density in Figure 7, Figures 2-6 and 2-7 show that the highest peak 1-hour levels and peak 8-hour levels are located in and immediately downwind of the major populated areas of the country, as expected. However, in the rural and 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.ppm. When compared to the population density map, the distribution of peak 8-hour ozone values across the country suggests a substantial contribution from man-made ozone on top of a significant background of ozone.

In contrast, the mean ozone concentration across the country, shown in Figure 10, 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 rural and remote 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 moderately-populated, highly-populated, rural, 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 personal 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 the mean exposures to a significant degree.

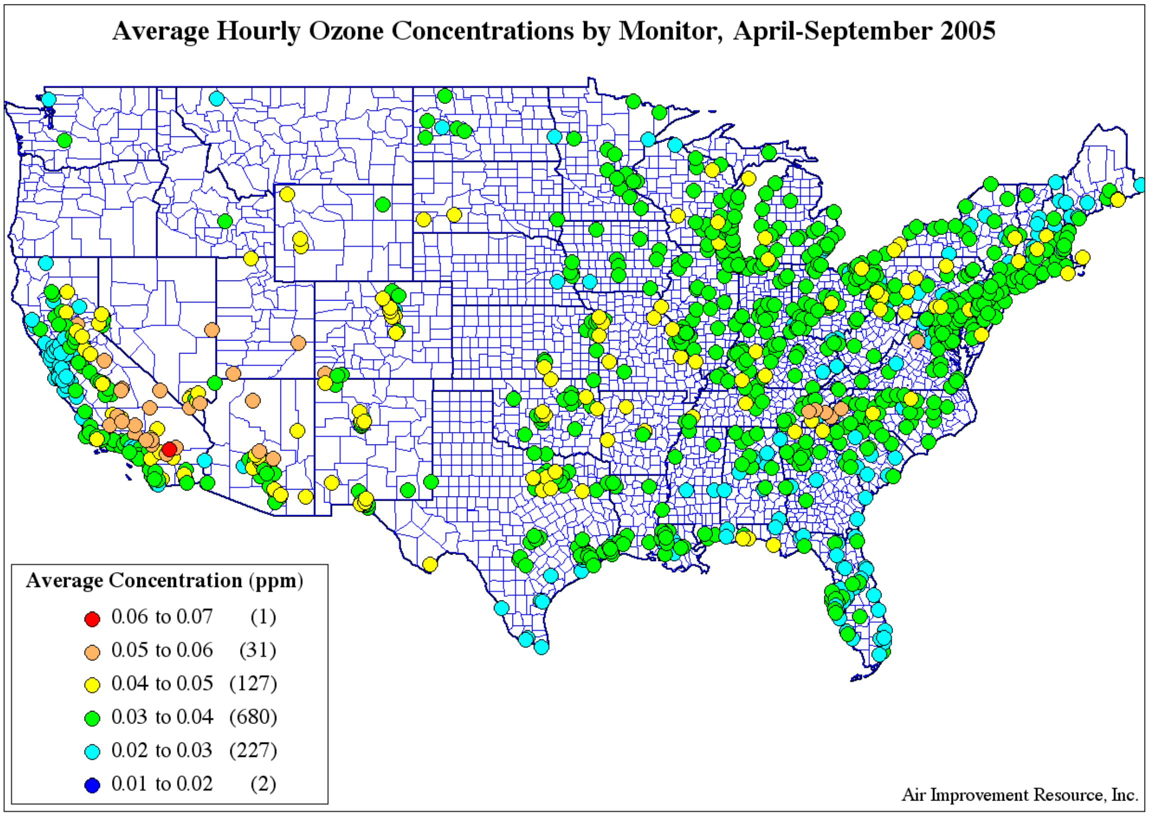


Figure 10

The second important real-world variance in ozone levels involves the impact of the nation's ozone control program. That program has yielded reductions of all three ozone precursors through national controls as well as local controls in major urban areas. . A comparison of the ozone trends in urban and rural locations provides additional evidence for the presence of a larger uncontrollable background than acknowledged in the SP. Although the text of the SP notes that that the trends in urban and rural locations are similar, a comparison of Figure 2-16 with 2-17 and Figure 2-18 with 2-19 of the SP shows that the upper percentiles of the peak concentrations have decreased substantially at urban and city-center sites but not at rural sites. In addition, the mean and lower percentile peak concentrations have not changed substantially at either urban or rural sites. CASAC pointed out that the trends at urban and rural sites are not similar. We would also point out that the massive nation-wide reductions in all three man-made precursors noted in the SP over the past twenty to thirty years have effectively reduced peak 1-hour and 8-hour concentrations in and immediately downwind of urban areas. However, they have not materially changed ozone in the more rural and remote areas of the country. Since the reduction in ozone in and immediately downwind of the major urban areas also resulted in lower levels of ozone and precursors transported further downwind, the lack of reductions in rural and remote areas suggests that non-U. S. man-made sources contribute substantially to the ozone concentrations in these areas.

The third real-world variance involves the reductions in emissions from weekdays to weekends. It has been clearly established that the reduced overall traffic and in particular reduced truck traffic on weekends results in a NOx-dominated reduction in emissions from weekdays to weekends. The ozone response to this emission change has been studied and reported by various investigators. In various areas of the country, this NOx-focused strategy has resulted in areas of ozone increase and areas of ozone decrease. This is consistent with known chemical behavior where a NOx reduction in a VOC-limited area will increase ozone and a NOx reduction in a NOx-limited area will decrease ozone. Heuss, et al. have published a national analysis of the weekend effect.[43] In the Western U. S. outside of major urban areas, there was no difference between weekday and weekend ozone. Since these rural and remote areas are NOx-limited, the lack of an ozone change in response to a 30 to 50 % reduction in ground-level man-made NOx emissions on weekends suggests that there is little local ozone production due to man-made ozone precursors over the rural and remote regions of the west.

In each of these three real-world instances, in which there are large spatial or temporal differences in U. S. man-made ozone precursor emissions, the response of measured ozone concentrations indicates the presence of an uncontrollable background with higher extremes than the SP and CASAC assume based on the GEOS-CHEM modeling.

The role of background in W126

The SP assumes, based on the GEOS-CHEM results, that background does not approach 0.040 ppm and, hence, that W126 is not influenced or confounded by concentrations thought to be background. This is an erroneousassumption. As noted in section II, the spatial pattern of W126 for 2001 is not consistent with contributions from only man-made emissions. In Figures 11 and 12, the three-month seasonal W126 (defined in the manner recommended for the secondary standard in the SP) is plotted in increments consistent with the recommendations of EPA staff and CASAC for 2003 and 2005. The monitoring sites denoted with open squares had too much missing data to calculate W126 following EPA's procedure. While there are an appreciable number of sites with W126 levels below the low end of the recommended range, there are also many rural and remote sites in the interior west and central plains with W126 levels above 7 ppm-hr. One reason for the lower W126 levels in the northern tier of western states is that the period of weather conducive to ozone formation is shorter in these states than in more southern states. For example, the ozone season in New Mexico is year round while the ozone season in Montana is only four months (June through september). Another factor may be the large quantities of monoterpene and sesquiterpene compounds that are emitted by pine trees in the northwest that react quickly with ozone.[44]

[43] J. Heuss, et al., J. Air & Waste Manage. Assoc., **53**, pages 772-788 (2003).
[44] D. Helmig et al., Environ. Sci. Technol., **41**, 1545-1553 (2007).

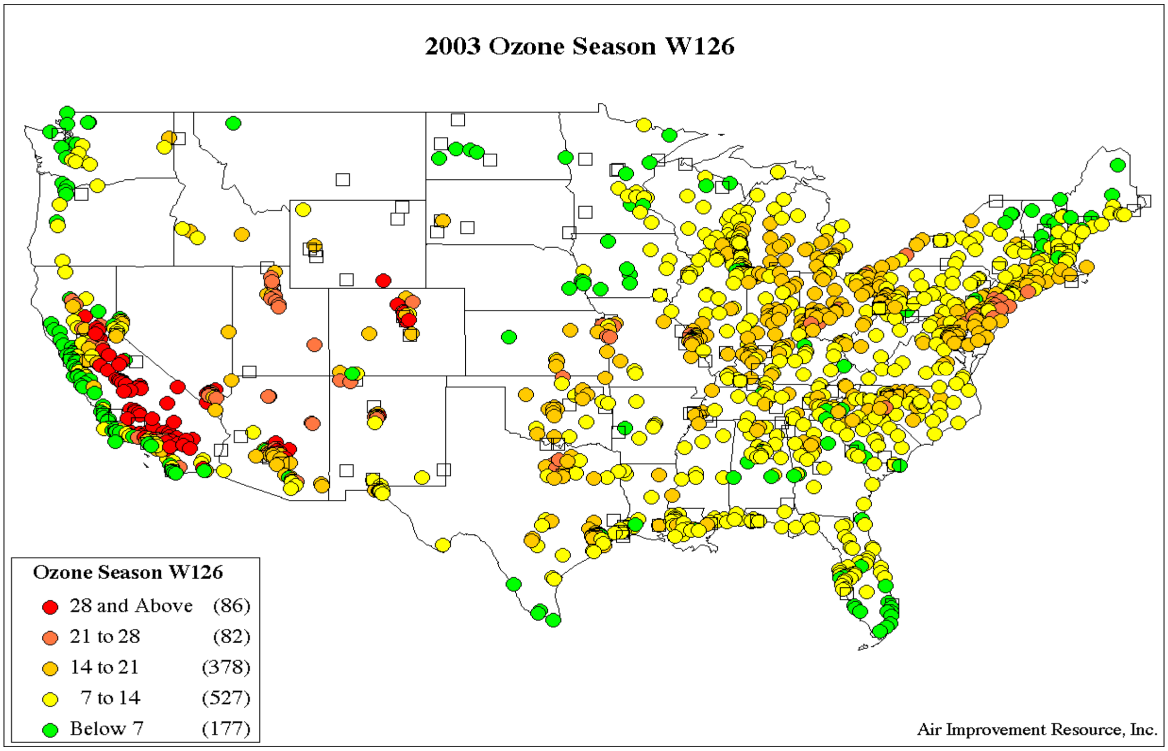


Figure 11 (monitoring sites denoted with open squares had too little data to calculate W126 following the EPA procedure)

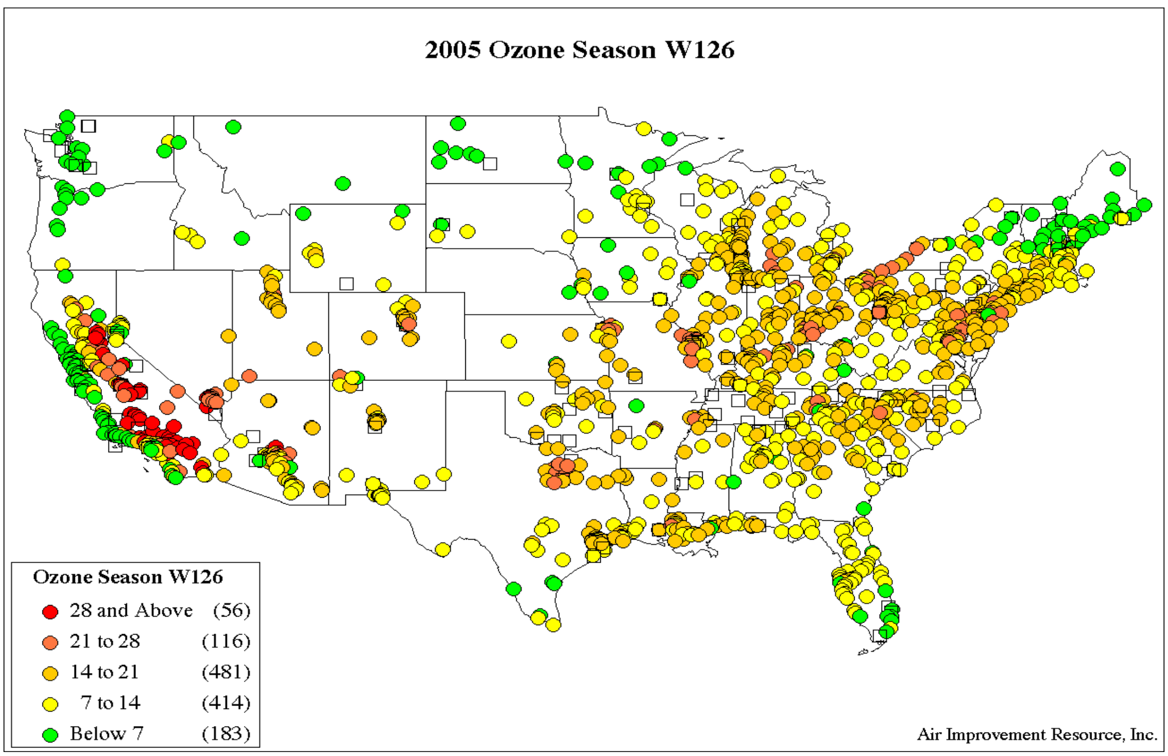


Figure 12

In Figure 2 in Section II, longer-term W126 data is plotted for a series of sites across the western U. S., from the California/Oregon border across Idaho, Wyoming and into South Dakota. All of these sites are substantially removed from major emission sources and have W126 values generally in the range of 0 to 16 ppm-hr. There are wide variations from year to year in W126 in Figure 2.

To evaluate a wider range of rural western monitors, long-term trends in W126 from rural monitors for western states east of California are plotted in Figure 13. To determine if a monitor (county) was "rural", the monitor had to be tagged as "rural", with a land-use of "agricultural", "desert" or "forest" or designated as being in a national park. The county could not be part of a MSA or CMSA. All but one of these sites exceeds the low end of the recommended range for a W126 standard and there is substantial variation 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 400 hours per year. In fact, almost all these sites had 8-hour design values of 0.06 or greater in 2002-2004. Second, the W126 metric varies substantially from year to year even in rural and remote areas that are not obviously influenced by anthropogenic sources. This means that if a W126 or other cumulative 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 W126 across the entire country indicates that W126 is partially confounded by background. A W126 standard as low as 7 ppm-hr would put a large portion of the country out of attainment, even locations that are far removed from any significant man-made precursor emissions. It would be impossible to develop rational implementation plans for such areas and, for other areas, complete elimination of man-made emissions might be required to demonstrate attainment. When the USEPA Administrator set the current federal 8-hour standard, she considered the influence of peak background levels and rejected alternative standards that were close to peak background levels that infrequently occur in some areas due to nonanthropogenic sources of O_3 precursors and would inappropriately target such sources.[45] In the current review, peak background has not been rigorously evaluated.

[45] 62 Fed. Reg. 38868, July 18, 1997.

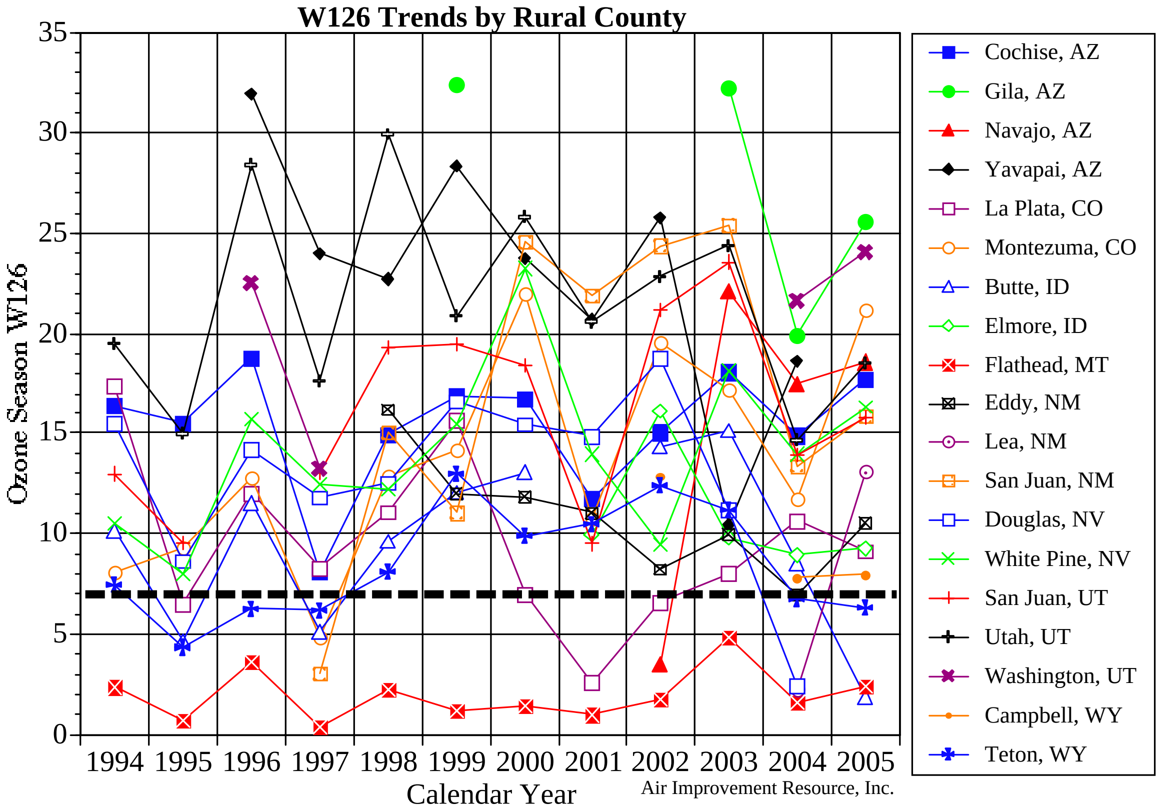


Figure 13

Summary

With an uncontrollable or Policy Relevant Background (PRB) that varies substantially, there will be times and places where the background approaches or exceeds the levels recommended by CASAC for both the primary and secondary standards. The SP attempts to limit the discussion of policy relevant background to the meteorological conditions conducive to peak urban ozone formation. While this is currently the limiting case for development of control plans, it would not be under the standards recommended by CASAC.

To illustrate, if the background is 0.040 ppm and the standard is 0.060 ppm, the amount of ozone that can be formed from man-made emissions is only 0.020 ppm. So even with a 0.040 ppm background, a 0.060 ppm standard would allow little room for human activities. On a day when the background is 0.060 ppm, there would be no margin for human activities. While this illustration over-simplifies the complex chemical and meteorological processes involved in ozone formation and transport, it demonstrates that transport of ozone from upwind natural and non-U. S. man-made sources can make some of the standards under consideration unattainable.

The SP's "simplistic" estimates of PRB downplay the true variability and complexity of the sources of background ozone that cannot be controlled by control of U. S. man-made emissions. Therefore, EPA must fully acknowledge the variability and develop estimates (using higher resolution models and observations) of peak background on the metrics being considered for primary and secondary standards. As noted above, when the CMAQ model was used to predict SUM06 and W126 in the Western U. S. for the year 2001, it failed miserably. Thus, available models cannot be used to evaluate uncontrollable ozone for the seasonal cumulative metrics in the western U. S. and observations must be used.

Appendix B Detailed Comments on Epidemiological Studies

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

Ozone is a well-known respiratory irritant and, for 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 EPA must address the question of whether ozone, in particular, is causally related to these endpoints.

The role of information from controlled studies

Chapter 6 of the SP fails to adequately discuss the importance of the neurally-mediated responses to ozone that are discussed in the CD. 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 that is particularly important in strong responders, 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 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 serious effects in controlled studies is an important consideration when it comes to evaluating the plausibility of the epidemiologic associations reported in the literature.

Information from systematic studies

The NMMAPS database is particularly important because, by including all the 90 largest U. S. cities with data, it avoids the issue of publication bias. The wide range of pollutant associations in the data, using the same methodology, demonstrates the inherent noise or variability in the data. The same wide pattern of associations ranging from strongly positive to strongly negative is observed in the individual NMMAPS data for all the pollutants studied at all the lags studied.[46] A comparison of the wide range of positive and negative associations in the robust NMMAPS data set compared to a truncated range of associations in the published literature as summarized by Stieb et al.[47] indicates that publication bias is a major issue in the air pollution epidemiology literature. The Stieb et al. meta-analysis of 109 mortality studies also demonstrates a remarkably similar pattern of associations for each pollutant in single-pollutant models that makes it difficult to implicate one pollutant over any other.

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 statistical issues with the General Additive Model were raised and many time series studies were re-analyzed, Ito[48] 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, there was a wide range of negative and positive risks in Detroit when all pollutants, lags, and endpoints were considered. Ito showed in separate figures that the wide range of associations occurred for each pollutant, including ozone. 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. Thus, time-series epidemioloogy of air pollution associations is a very blunt tool. Although CASAC raised this issue in its June 2006 letter, the panel's recommendations essentially accepted positive associations as evidence of health effects.

If all positive associations are assumed or concluded to be health effects then all negative associations should be considered health benefits due to ozone or other air pollutants. Since no-one believes that ozone or other air pollutants actually provide health benefits on a day-to-day basis, the assumption that all positive associations are health effects is not plausible. For example, the various ozone /mortality associations in Figure 7-17 of the CD would imply that ozone is dangerous in about a third of cities but is harmless or even beneficial in about half the cities. This is not biologically possible, if it is assumed that these associations represent causal relationships.

[46] 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.
[47] Stieb et al., J. Air & Waste Management Association, **52**, 470-484, 2002; Stieb et al., J. Air & Waste Management Association, **53**, 258-261, 2003.
[48] K. Ito, pages 143-156 in Health Effects Institute, Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, May 5, 2003.

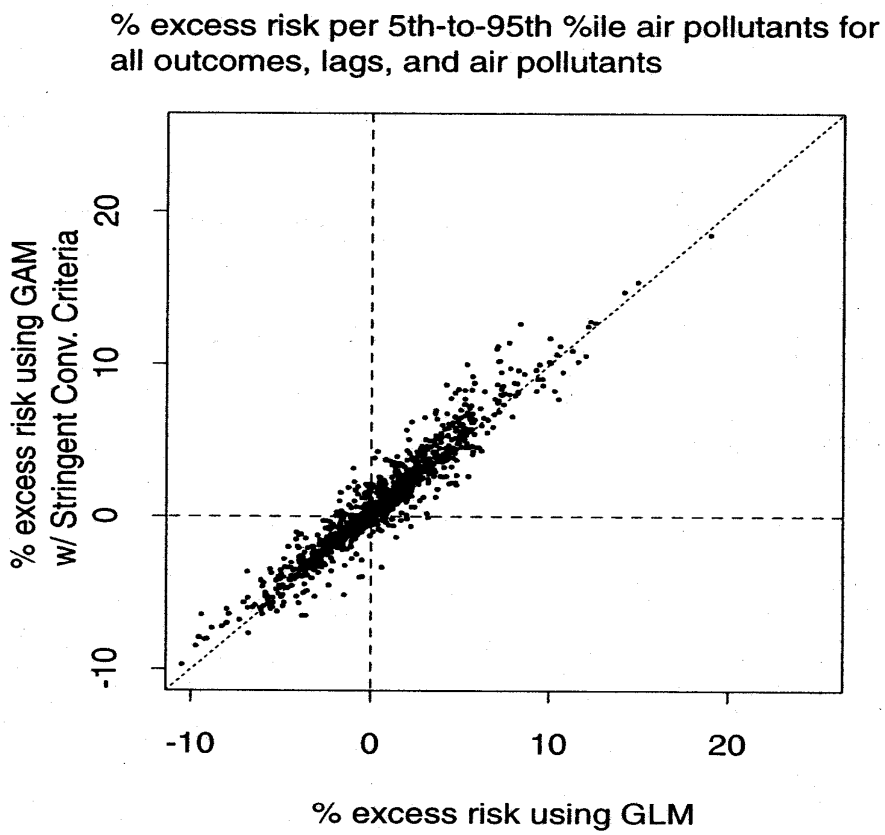


Figure 14 - Figure 2 from Ito 2003

The second draft SP acknowledged that "the most important uncertainty is the extent to which the associations between ozone and the health endpoints included in the assessment actually reflect causal relationships."[49] Although there clearly are positive associations in the literature for the health endpoints used in the risk assessment, the overall pattern of associations in systematic studies in the literature is not biologically plausible or consistent with causality. In multi-city studies, there is an implausibly wide range of associations for mortality 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.

[49] Second Draft Staff Paper at page 5-76.

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)

NMMAPSII (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 as shown in CD Figure 7-17
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 associations.

The range of associations in the published literature

The SP fails to fully acknowledge that the body of studies since 1996 demonstrates major uncertainties and issues related to model selection and publication bias. These issues are discussed in the next two sub-sections. 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. This is reflected in CASAC's June 5, 2006 letter to the Administrator that raises serious issues regarding the use of time series studies.

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 <u>and</u> 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 %

Publication bias

As numerous weak but positive associations of various air pollutants with serious health outcomes have appeared in the literature, there is increasing concern over the issue of publication bias. Such associations have been prominent factors in the recent review of PM and ozone standards, and they are now an issue with the upcoming NO_2 and SO_2 reviews. For the ozone review, the EPA commissioned three new meta-analyses of ozone/mortality associations. When these studies were published they were accompanied by two commentaries. 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."[50] He notes a major discrepancy between the estimated associations from the comprehensive National Mortality and Morbidity Air Pollution Study (NMMAPS) that evaluated the 90 largest U. S. cities and the meta-analyses. The discrepancy raises the issue of publication bias. Goodman cautions that "depending on published single-estimate, single-site analyses is an invitation to bias."[51] 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."[52]

Model selection uncertainty

An important systematic analysis was carried out by Koop and Tole[53] who used Bayesian model averaging to evaluate model uncertainty in time series analyses using an extensive set of pollutants 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.[54]

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

[50] S. Goodman, Epidemiology , **16**, pages 430-435 (2005) at page 430.
[51] Id. at page 430.
[52] Id. at page 431.
[53] G. Koop and L. Tole, J. of Environmental Economics and Management, **47**, pages 30-54, 2004.
[54] Id. at pages 46 and 47.

"…may lead researchers to make misleading inferences about pollution-mortality effects, thereby seriously underestimating the true uncertainty in the statistical evidence."[55]

In 2003, when issues were raised with the statistical model used to analyze many time series data sets during the review of the particulate matter standards, many time series studies were re-analyzed. The Health Effects Institute Special Panel that reviewed the re-analyses concluded "…neither the appropriate degree of control for time in these time-series analyses, nor the appropriate specification of the effects of weather, has been determined."[56] They went on to indicate that "this awareness introduces an element of uncertainty into the time-series studies that has not been widely appreciated previously." In fact, the Koop and Tole analysis is the kind of analysis the HEI 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.

Koop and Tole, as noted above, show that individual model results are unreliable. During the federal particulate matter review, AIR presented evidence that led us to the same conclusion.[57] By empirically comparing the results of different time series studies of the same city by different investigators, AIR 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. During the California review of the state ozone standard, the Air Resources Board staff made the same point noting that "alternative analyses of data from the same city sometimes resulted in differing results".[58] There are, in fact, many examples in the literature of this phenomenon. It is a practical example of the model selection issue that has been raised in the HEI Special Report on re-analysis of time-series studies.

Given the issues of publication bias, model selection uncertainty, and the wide range of both positive and negative associations in systematic analyses, the only conclusion that can be drawn is that, while there are many positive epidemiological associations in the literature, individual city studies are not reliable. This conclusion may be considered controversial. It contrasts with the view that any statistically significant positive association is likely real and causal. However, many competent investigators are becoming more skeptical about the interpretation of weak air pollution associations.

For example, Moolgavkar has expressed severe reservations. He published several studies of mortality and hospital admissions in Los Angeles. In one of his studies, Moolgavkar noted discrepancies between his findings and that of other studies in Los Angeles. Moolgavkar acknowledged that he did not know how to explain the discrepancies. Although there were differences in the methods of analysis, Moolgavkar at

[55] Id. at page 40.
[56] Health Effects Institute, Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, May 5, 2003, at page 269.
[57] See AIR, Inc. comments on 2nd, 3rd, and 4th drafts of USEPA PM Criteria Document.
[58] California Environmental Protection Agency, June 21, 2004 Public Review Draft for Ozone Standard at page 12-76.

the time did not think the differing statistical analyses could explain the discrepancies. However, he went on to indicate that "If indeed they do, then one must conclude that results of time series analyses can be quite sensitive to statistical approaches."[59]

In the re-analysis of his results in 2003, Moolgavkar reported that some results changed and some did not. Based on his results and those of other investigators, he concluded, "given that different analytical strategies can substantially affect the estimates of effects of individual pollutants, I believe that no numerical estimates are very meaningful."[60]

Vedal and co-workers[61] have also expressed the concern that pollutant/health associations may not be effects of the pollutants themselves, but rather of some other factors present in the air pollution-meteorology mix.

In addition, Lumley and Sheppard point out 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." [62]

Conclusions regarding causality

The current practice of using central station monitoring data, central station weather data, and available health statistics yields many weak positive associations for various pollutants, including ozone. However, it is known that the methodology is subject to problems of measurement error and exposure miss-classification as well as severe collinearity between weather and pollution variables. When the uncertainty due to model selection issues is added, and the potential for publication bias is considered, the interpretation of a subset of positive findings as causal becomes problematic.

Nevertheless, EPA staff indicates that there is clear and very strong evidence supporting the judgement that the relationships are causal for respiratory symptoms in asthmatic children and respiratory hospital admissions. Staff acknowledges that there is greater uncertainty regarding mortality, describing the overall body of evidence as highly suggestive.[63] We believe the underlying data is much less consistent than portrayed in the SP and that staff's interpretation is problematic since it fails to consider the uncertainties noted above. In the following sub-sections, we discuss the data for the major endpoints under consideration.

Respiratory symptoms in asthmatics

The SP refers to a number of studies that report robust associations of ozone with daily symptoms/asthma medication use. 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

---

[59] S. Moolgavkar, Environmental Health Perspectives, **108**, pages 777-784 (2000) at page 781.
[60] S. Moolgavkar, in Health Effects Institute, Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, May 5, 2003, at page 198.
[61] S. Vedal et al., Environmental Health Perspectives, **111**, 45-51, 2003.
[62] T. Lumley and L. Sheppard, Epidemiology, **14**, 13-14, 2003.
[63] Staff Paper, *supra* note 1, at page 5-93.

the U. S. The SP quotes the CD as indicating that the panel studies as a group indicate a positive association between ambient concentrations and respiratory symptoms and increased medication use in asthmatics.[64] The evidence is noted as much stronger than in the previous review.

As noted in AIR's previous comments, the reliance on single-pollutant results from Mortimer et al. 2002 is 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_2 was measured in all eight cities, and when SO_2 was included in two pollutant models, the ozone coefficient was reduced from 1.16 to 1.11 and became insignificant. For seven areas both SO_2 and NO_2 were available. In multi-pollutant models the ozone coefficient was reduced to 1.06 and again was insignificant. 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_2 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_2 in single-pollutant models and with SO_2 in multi-pollutant models. In fact, the authors of the study specifically focus on ozone as a marker of summer air pollution and refer to the study's finding as effects of summer-time air pollution rather than ozone, per se.

A new study by Schildcrout et al. 2006[65] evaluated the relation between various pollutants and asthma exacerbations (daily symptoms and use of rescue inhalers) among 990 children in eight U. S. cities and reported associations with several pollutants but ozone was unrelated to exacerbations. When the new Schildcrout et al study is considered along with the nine other studies that are described in the CD as having limited or a lack of evidence for ozone effects, the overall findings relative to effects of ozone on asthmatics on symptoms and medication use in the SP should be described as mixed and inconsistent rather than clear and very strong evidence of an ozone effect. Confounding by bioaerosols may be 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.

Emergency department visits and hospital admissions

There is a very wide range of single-pollutant ozone 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

[64] Staff Paper, *supra* note 1, at page 3-12.
[65] J. Schildcrout; L. Sheppard; T. Lumley; J. Slaughter; J. Koenig; G. Shapiro, Ambient air pollution and asthma exacerbations in children: An eight-city analysis, Am. J. Epidemiology, **164**, pages 505-517 (2006).

admissions, the range is from – 15 % to + 15 % per standard ozone increment as shown in Figure 7-9. The interpretation of this data has changed during the ozone review In the second draft CD and the first draft SP, the conclusion was drawn that the body of evidence remains inconclusive regarding effects of ozone on the risk of emergency department visits.[66] However, that conclusion was removed from the final CD and the SP and replaced with a statement indicating that there are generally positive associations in summer analyses for asthma ED visits.

A number of rigorous evaluations of this literature come to different conclusions. Anderson, et al. 1998 reported that ozone, SO_2, NO_2, and particles all had positive associations with emergency asthma admissions in London, but that there was a lack of consistency across age groups and seasons.[67] Anderson et al. also evaluated 15 other studies of air pollution and daily asthma emergency room visits or admissions in the literature. They evaluated the consistency of these studies and report that, in the all-age group, three studies did not find significant associations with any of the pollutant assessed and the proportions with significant findings for ozone, SO_2, NO_2, and particles were 7/14, 6/12, 2/9, and 7/15, respectively. Similar results were found for adults and children considered separately. They conclude "Taken overall, it is apparent that the evidence is not coherent as to whether there is an effect of pollution or the responsible pollutant." They go on to indicate that ozone, SO_2, and particles were significant in no more than half the studies and that only about a quarter of the studies found significant effects for NO_2. They list a number of possible reasons for the lack of consistency, including false negatives due to lack of statistical power and false positives due to chance, multiple significance testing, post hoc hypothesis testing, or publication bias. They also note differences in pollution level and mix between cities, the presence of highly correlated pollutants, and that pollutants acting as surrogates for unmeasured pollutants or ambient aeroallergens may be involved. They conclude that, while there is evidence that all of the pollutants may have an effect on asthma, there is a lack of consistency in the specific pollutant responsible.

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 first draft SP. The California review[68] 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

[66] See first draft SP at page 3-8 which refers to page 7-62 of the second draft CD concluding that "the body of evidence remains inconclusive regarding the effects of ozone on the risk of emergency department visits."
[67] Anderson et al. *supra* note 9.
[68] California Air Resources Board Staff Report, Review of the California Ambient Air Quality Standard for Ozone, March 11, 2005,

acknowledged that there is no clear pattern of association with specific respiratory disease outcomes.[69]

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 strong evidence of effects.

Increased school absences

The SP notes that the association between school absences and ozone was assessed in tworelatively large field studies.[70] The SP discusses the Chen et al. 2000 and Gilliland et al. 2001 studies of school absences, but quotes the CD that further replication is needed before firm conclusions can be drawn. The final CD, however, discusses three analyses of the Southern California Children's Health Study data that come to very different conclusions. In addition, the final CD deleted discussion of the Park et al. study presumably because it violates the temporality requirement. Overall, the school absence data is weaker than noted in the SP.

Premature mortality

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[71] that extends the previous analyses, and the Huang et al. 2005 analysis of a subset of 19 large U. S. cities. AIR provided detailed comments on these studies in previous submissions.

In particular, we noted major deficiencies in the Bell et al. 2004 paper that omitted winter ozone results and limited the analyses regarding other pollutants. We also noted that if the reported association is causal, it 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 on a day-to-day basis.

---

[69] California Air Resources Board Draft Staff Report, Review of the California Ambient Air Quality Standard for Ozone, June 2004.
[70] Staff Paper, *supra* note 1, at page 3-12.
[71] M. L. Bell, et al., "Ozone and Short-term Mortality in 95 US Urban Communities, 1987-2000," JAMA, **292**, 2372-2378, 2004.

We also showed how 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 lower ozone levels in the cities were associated with higher mortality associations. 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. 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 were not rigorously considered in the development of the staff recommendations. Instead, the staff listed a number of reasons why the ozone association might be causal and did not list the several reasons why it might not. The staff then, without sufficient justification, changed the interpretation of "highly-suggestive" of causality to "likely" causal.

The SP 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. 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. 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 was not 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.

In the HEI re-analysis, in single pollutant models, ozone had a small positive combined association in summer but also a small negative combined association in winter.[72] In multi-pollutant models for the summer case, the combined positive ozone association was

[72] Dominici, F.; McDermott, A.; Daniels, M.; Zeger, S.; Samet, J.; Revised analyses of the National Morbidity, Mortality, and Air Pollution Study: Mortality among residents of 90 cities, *J. Toxicology and Environmental Health, Part A*, **2005**, *68,* pages 1071-1092.

reduced and was not significant at all three lags evaluated as shown in Figure 7-31 of the CD. Ozone is not the only pollutant for which NMMAPS seasonal results have been published to date; a seasonal PM10 analysis has also been carried out.[73] Peng et al. 2005 used updated mortality data from 1987-2000 in 100 cities. Summer was the only season for which the combined PM10 effect was statistically significant. An analysis by geographic 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. The authors note several possible explanations. One obvious hypothesis is that the most toxic particles have a spring/summer maximum and are more prevalent in the Northeast. Another possibility noted by the authors is that there could be a seasonally varying bias that is yet unidentified.

The NMMAPS database is valuable because, by including all the 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, CO, NO_2, and SO_2) at all the lags studied.[74]

The SP and CD dismiss concern over confounding by gases since the peaks of the various pollutants occur in different seasons and because ozone is generally not highly correlated with other pollutants. 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 effects related to changes in weather and human behavior. Therefore, confounding by weather is a candidate.

A recent paper by Keatinge and Donaldson[75] 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 –

[73] Peng, R. D.; Dominici, F.; Pastor-Barriuso, R.; Zeger, S. L.; Samet, J. M.; Seasonal analyses of air pollution and mortality in 100 U. S. Cities, *Am. J. Epidemiol.*, **2005**, *161*, 585-594..
[74] 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.
[75] W. R. Keatinge and G. C. Donaldson, “Heat acclimatization and sunshine cause false indications of mortality due to ozone, *Environmental Research*, **2006**, *100*, 387-393.

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_2, 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. 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 considered in the interpretation of the risk assessment.

As noted above, the SP 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.[76] 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." Although the SP mentions the issue of publication bias, it is discounted by noting that ozone-mortality associations remained after accounting for the bias.

Comments on the risk assessment for epidemiological endpoints

There are three important issues or sources of uncertainty noted in the SP. The first and main issue in interpreting the epidemiologic associations, as discussed in detail above, is whether they are real health effects or not.

A second major uncertainty is related to the adequacy of ambient monitors as surrogates for population exposure.[77] There are two points here. One is that in many studies there is little or no correlation between ambient measurements and personal exposures, particularly for populations that spend the bulk of their time indoors. Second, the fact that indoor exposures are reduced by a factor of two to four compared to ambient measurements 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. For example, Figure 5-15 of the SP shows that the bulk

[76] Goodman, S. N.; The Methodologic Ozone Effect, *Epidemiology*, **2005**, *16*, 430-435.
[77] Staff Paper, *supra* note 1, at page 5-43.

of the estimated mortality risk comes from low ambient exposures and it is known that the personal exposures of the frail population are much below ambient levels.

A third major area of uncertainty noted by staff is a set of issues related to the empirically-estimated concentration-response relationships.[78] 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. 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 being outside at the time of elevated ozone and undergoing 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. Earlier drafts of the CD and SP 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.

As a result of these uncertainties, the risk results presented in Chapter 5 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. While there are positive associations in the literature for all the endpoints being considered, such associations occur for all the pollutants that have been evaluated, and the full range of associations 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[79] about the use of point estimates from time-series to set regulatory standards need to be carefully considered by CASAC and EPA.

[78] Staff Paper, *supra* note 1, at pages 5-42.
[79] Koop, G.; Tole, L.; Measuring the Health Effects of Air Pollution: to What Extent Can We Really Say that People are Dying from Bad Air, J. *of Environmental Economics and Management*, **2004**, *47*, 30-54.