

Detailed ExxonMobil Comments on the 1st Draft Staff Paper for Ozone

Summary

- As clearly noted by CASAC, the existing data are inadequate to establish a causal relationship between ambient ozone exposure and mortality. Therefore, it is not appropriate for EPA to quantify acute ozone mortality.
- The existing concentration response functions for acute ozone mortality are unreliable and should not be used for risk assessment.
- The linear-no-threshold approach to assess acute mortality is not supported by the available scientific data. This approach markedly exaggerates the risks and will not result in an accurate and reliable risk assessment.
- EPA has not considered the key uncertainties in their draft risk assessment. EPA must present a risk assessment case that does not include acute mortality, and additional cases where mortality is not assessed below plausible cut points derived from human clinical and toxicology data.

Introduction

In the following Detailed Comments section we address the important points raised in the Staff Paper using the same numbering convention as the Staff Paper. For the sections noted, we first state the position as taken in the Staff Paper then follow with our reasoning and interpretation of the data. Our interpretations are often repeated in several sections because the Staff Paper often deals with the same, or similar, points in several sections.

Detailed Comments

Section 3.4. Integrative Assessment of the Evidence from Epidemiological Studies

In this section, EPA provides a Hill oriented discussion on whether or not the existing evidence supports a causal relationship for the acute ozone mortality hypotheses. As we have indicated in our detailed comments on the draft CD, the data do not support a causal relationship. Briefly, we present below our concerns with the Hill assessment as presented in the draft Staff paper.

3.4.1 Strength of Associations

EPA Staff Paper Position

- Effect estimates between O₃ and various health outcomes are generally small in size. The magnitude of these associations, while small, is found to be relatively consistent between studies.
- In considering both the magnitude and statistical strength of the associations, a pattern of positive and often statistically significant associations can be seen between mortality and respiratory morbidity and short term exposure to ozone.

ExxonMobil Comments on EPA Staff Paper Position

This guideline is an estimate of the magnitude of the association. The stronger the association (the greater the magnitude of the risk ratio between exposed and non-exposed) the more likely the association is causal. The effects of exposure should be greater than effects of hidden uncertainty or the risk estimates will be inaccurate because of probable confounding and bias. A rough guideline in traditional epidemiology is that relative risks >1.5-2 fold are necessary to suggest causal association. Relative risks less than this are likely due to confounding. But a risk could be low and still infer causality if other guidelines are satisfied.

Statistical significance and statistical power of effects estimates are not considerations about the *strength* guideline. They are considerations about whether the results could be due to chance or random error. These items are discussed later in a separate discussion on *robustness*. Similarly, *consistency* is not related to the strength of association and should be not discussed in this section; it is addressed later in a separate section. Therefore, EPA should remove the discussions on robustness and consistency from this section.

The relative risks of the ozone mortality studies are in the range of 1.000 -1.008 and are therefore *very weak*. Associations of less than 1% are particularly susceptible to hidden uncertainties of confounding and bias. For this to be an accurate estimate of an effect this small, there must be essentially complete control of confounding and bias. However, complete control is not possible.

For example, the CD notes on p 8-8 that exposure data “should provide a relative assignment of exposure with time, if: concentrations are relatively uniform across the region; time-activities pattern are roughly the same across the study population; and housing characteristics...are relatively constant for the study area.” But these if’s are not well met as Bates (Bates 2005) notes that O₃ concentrations are lower in the inner city than in the suburbs because of NO from vehicles that “quenches” O₃, forming NO₂ and reducing O₃. Concentrations are not uniform in distribution across the city or across time (peak 12-3 pm) and air conditioning reduces ambient concentrations ~85%. Levy et al (Levy et al. 2005) noted that as air conditioning prevalence increased from 0% to 60%, O₃-associated mortality/10 ppb 24-hour average decreased from about 1.015 to 1.000.

The 1.008 risk estimate from the meta analyses is biased high due to probable “publication bias” as well as other biases. Adjustments for publication bias assessed by multi-city analysis suggest an overall effects estimate is about 1.0025, or about 1/3 the quoted estimate of 1.008. If lags of 1 or 2 days were used, the overall estimate would be reduced to less than 1.002 (Bell et al. 2005).

Ito et al (Ito et al. 2005) reported several effect estimates besides 1.008. For example, 1.006 was an estimated effect level when PM was in the model. Selection of a different weather model could produce a 2-fold change in estimate, which reduced the effect estimate to about 1.004 or 1.003 if the model was the one used in the NMMAPS analysis.

Levy et al (Levy et al. 2005) reported that using 1-2 day lags instead of a 0-day lag reduced the effect estimate to about 1.004. Adjustments for other biases such as publication bias, a different weather model, and other pollutants would likely reduce this estimate even further.

In conclusion, the associations between ambient O₃ concentrations and acute mortality are *very weak* indicating this guideline does not support a causal association. If important factors such as lags, weather model, and publication bias are considered the associations become even weaker. It is reasonable to conclude that precision of an estimate in epidemiology studies (especially ecological studies) is something less than 3 and perhaps 2 significant figures. Assuming epidemiology results cannot ensure the precision of an estimate beyond 1 or 2 significant figures, these single point effect estimates suggest no association as the rounded-off estimates are 1.0 or 1.00.

3.4.2. Robustness of Associations

Robustness of the associations is the second guideline mentioned in the Staff Paper. In the CD, this is defined as “evaluating the impact of alternative models, model specifications for temporal trends and meteorological factors, and potential confounding by co pollutants.” Also issues related to exposure assessment and measurement error are considered.

EPA Staff Paper Position

- Ambient concentrations serve as valid surrogate measures for aggregate personal exposure in time-series studies. The ambient concentrations generally over-estimate true personal exposure so the risk estimate is biased toward the null.
- Adjusting for temporal trends and meteorological factors is said to be critical. Confounding for seasonal variability is “controlled effectively by stratifying...by season.”
- Assessment of confounding by copollutants indicates adjustments for PM do not substantially change the risk estimates.

ExxonMobil Comments on EPA Staff Paper Position

EPA’s own data, presented at the recent CASAC meeting, indicate that there is a very low correlation between ambient and personal ozone exposure, and the correlation varies across cities. For example, the correlation was less than 20% in Boston and less than 10% in Baltimore. The published literature supports the low correlation. This invalidates EPA’s key assumption that ambient measurements are a valid surrogate for aggregate personal exposure.

We suggest that robustness as a guideline should consider two types of uncertainty: *statistical uncertainty* and *hidden uncertainty*. *Statistical uncertainty* is an evaluation of whether the findings are due to chance and so is largely dependent on sample size. The CD mentioned statistical significance and statistical power under Strength of Association. But the magnitude of the effect estimate is independent of statistical significance, and both significance and sample size are more appropriately considered as a separate guideline for evaluating results.

Most of the single-city studies in the meta-analyses are not statistically significant although the combined results are significant. Also, the power of O₃ air pollution studies can be somewhat misleading. If the sample size is very large then minuscule risk estimates of no plausible biological significance are considered important and overpower hidden uncertainties. As an example of minuscule risk and lack of biological significance, we see that the upper or lower confidence interval is sometimes the same as the risk estimate as in the 95-city study where the rounded-off risk estimate with 95% confidence interval is 1.01 (1.00-1.01).

The importance of statistical significance is over-emphasized. Reporting only confidence intervals implies random sampling errors are the only uncertainties worth measuring or reporting. The p-value does not provide an answer for a causal question. Statistical significance is neither necessary nor sufficient for inferring causality and does not supply the final word. Over-emphasis on statistical significance should be avoided as systemic error is usually greater than random error. As Hill noted "the glitter of the t table diverts attention from the inadequacies of the fare." The more important question is the "fare," including evaluation of hidden uncertainties of measurement error, confounding, bias, etc. Statistical precision (small p-value or narrow confidence interval) should not be equated with validity.

Hidden uncertainties include such things as confounding from co-pollutants, model choice (e.g., effect of smoothing in weather models, effect of different lags), publication bias from choosing the largest effect in single studies versus less subjectivity in model choice in multi-city studies, etc. These hidden uncertainties are more important than a precise statistical test, especially when sources of error are of greater moment than random error (Phillips et al. 2004). It is still not possible to accurately quantify the hidden uncertainties. But the series of meta-analyses on O3 mortality and some of the lessons from PM time-series studies indicate important hidden limitations that the CD has not taken into account.

Bates (Bates 2005) noted the high correlations of heat and O3 in the summer which makes it difficult to attribute individual effects to one or the other exposure. Ito et al make these limitations more explicit. Weather adjustment models alone can change effect estimates by a factor of 2. Models with smoothers fit the data better but have greater concavity so "model validations...are needed." "Statistical properties aside, it is not clear whether these models actually adjust for weather effects. Because daily fluctuations of weather and air pollution are related, it is possible that these models may ascribe ...some of the real pollution effects to weather." Therefore model uncertainties should be taken into account (Ito et al. 2005).

Goodman (Goodman 2005) noted that risk estimates from single-city studies are biased upward about 3-fold because of cherry-picking the strongest signal. Meta-analysis is subject to bias as shown by the elevated effect estimates of single-city combined results relative to multi-city studies such as NMMAPS (Goodman 2005) and indicates the presence of "publication bias" (Bell et al. 2005).

Adjustments for weather may not be appropriate as most studies used the same regression designs as were used in PM studies, but O3 is more highly correlated with temperature than is PM (Ito et al. 2005).

There is no objective or feasible method for selecting the appropriate model or amount of smoothing to adjust for weather (HEI Committee analysis of revised analysis of PM time-series studies). Ito et al (Ito et al. 2005) noted differences of 1.1% between 4 different weather models and 0.4% differences between smoothing models with 12 df/y and 26 df/y. These hidden uncertainties on model choice are greater than the effect estimates.

Meta-analysis has limited interpretability because of differences in analytical methods and because most studies have not adequately addressed influential factors (~90% are not applicable to use for regulations because substantial uncertainties remain) (Levy et al. 2005).

Single-city studies must be used with "caution" or not at all because they have about a 3-fold upward bias due to model choice (Bell et al. 2005). Ito et al (Ito et al. 2005) noted the large influence of model choice on effect estimates by drawing attention to the "markedly different estimates" in the 3 cities analyzed by different authors. In Amsterdam, London and Santiago the associations were positive for one set of authors and negative for the other set of authors.

3.4.2.1 Exposure Error

EPA Staff Paper Position

- Ambient ozone concentrations provide a valid surrogate measure for aggregate personal exposure in time series studies.

ExxonMobil Comments on Staff Paper Position

Due to the very poor relationship between ambient and personal ozone exposure, the time-series methodology is of questionable value for assessing the health effects of ozone. This is confirmed by data presented by EPA at the recent CASAC meeting, where a very poor correlation was reported between ambient and personal ozone in both Baltimore and Boston. Ambient ozone concentrations are often lower in the inner city than suburbs because nitrogen oxides from vehicles “quenche” ozone, forming nitrogen dioxide and thereby reducing ozone. Ozone concentrations outdoors are not uniform in distribution across city or time, as peak levels appear from 12-3 p.m when ozone is also very reactive. When ozone enters indoor environments, it reacts with surfaces and is largely destroyed. Air conditioning also decreased indoor levels. As a result, the levels measured on outdoor monitors are a very poor, and are a nearly meaningless surrogate for personal ozone exposure. Given these problems, it is not possible to ascertain whether the results of ozone time-series studies overestimate the effects of ozone, as alleged by EPA, or underestimate the health effects. Rather, given the extremely poor personal to ambient relationship and resulting exposure error, it is not possible to use results of time-series studies to accurately assess the health effects of ozone. Certainly, as noted by Brauer et al., 2002, it is not possible to use these studies to determine reliable concentration response relationships, particularly at low levels where the presence of other pollutants will confound the findings for ozone. In particular, fine particulate matter, which has been shown to penetrate indoors and therefore demonstrates a higher ambient to personal correlation, may be confounding any effect that ozone might have at low exposure concentrations.

3.4.2.2 Confounding by Copollutants

EPA Staff Paper Position

- Ozone effect estimates are robust to and independent of the effect of other pollutants.

ExxonMobil Comments on the EPA Staff Paper Position

First, very few of the existing time series studies for ozone have included PM_{2.5} as a potential confounder. Most of the existing studies have evaluated potential confounding effects of “PM” using PM₁₀. It is well known that a much higher correlation is observed for fine particulate matter and ozone versus coarse PM, particularly in summer when ozone risk coefficients for mortality are higher. Levy et al., one of the few authors attempting to evaluate potential confounding by fine PM, indicated that much more work is needed to explore potential confounding of the ozone mortality relationship with fine PM. Hueng et al. observed significant confounding of the ozone mortality relationship or PM. Therefore, overall, it is entirely possible that much of the acute mortality risk attributed to ozone is due to fine PM. In any case, it is clear that the Hill criteria for adequate assessment of potential confounders has not been met.

3.4.2.3 Model Specification

EPA Staff Paper Position

- Ozone effects on various health outcomes are robust to various model specifications.

ExxonMobil Comments on Staff Paper Position

The results of the recent meta analyses have confirmed the existence of model dependence in the results of ozone time-series studies.

3.4.3 Consistency

As described in the staff paper, *consistency* relates to the persistence of an association between exposure and outcome in multiple studies of adequate power in different person, places, circumstances and times.

EPA Staff Paper Position

- The Staff Paper states that the magnitude of the effect estimates is relatively consistent across the results of recently published studies of associations between short-term ozone exposure and mortality.

ExxonMobil Comments on Staff Paper Position

There is a high degree of heterogeneity for cities reported by different authors. There is also a high degree of heterogeneity across cities in multi-city studies.

Ito et al (Ito et al. 2005) noted the large city-to-city variation in the cities included in their analyses. They report a large index of heterogeneity of 77% that would preclude any combining of the data into one final number as was done in their paper. The range of the estimated percent excess deaths was from approximately –9% to +9%. A similarly high degree of heterogeneity across cities was also observed in the study by Huang et al. (see Table 1). Ito et al's comment in the discussion that their results "are fairly consistent with results from other recent meta-analyses" ignores the important condition that the meta-analytic results are based on inappropriate, or heterogeneous, data. Other meta-analyses also suffer from the same condition as seen in the results of Huang, et al, (2005) shown in Table 1.

Ito et al (Ito et al. 2005) noted the large influence of model choice on effect estimates by drawing attention to the "markedly different estimates" in the 3 cities analyzed by different authors. In Amsterdam, London and Santiago the associations were positive for one set of authors and negative for the other set of authors. Another example of an inconsistent pattern for individual cities reported by different authors is illustrated by comparing the results from Bell et al. and Huang et al. (see Table 1). Although Huang, et al considered cardiovascular and respiratory mortality (per 10 ppb increase in ozone during June through September) and Bell, et al considered mortality (per 10 ppb increase in ozone during the full year), it would be expected that the relative rankings among the 19 cities would be much closer than they are (the rank correlation is 0.12).

Goodman (2005) noted that “depending on the study published, single-estimate single-site analyses is an invitation to bias.” The most plausible explanation for the bias is that investigators “tend to report...the analysis that produces the strongest signal” and make model choices that positively affect the sign and magnitude of the signal. Thus a consistent finding of positive associations may be due to a consistent choice of cherry-picking models to produce positive findings. This is possible because there are “innumerable model choices.” Levy et al (Levy et al. 2005) noted that there were multiple studies where ozone was statistically insignificant without reporting the results, presumably because “such findings may be less likely to be published.”

The Staff Paper makes the point that mostly positive studies with the same magnitude of effect estimate provide consistency. Bell et al (Bell et al. 2005) note that the hypothesis of homogeneity was rejected for both US cities and all cities. Ito et al (Ito et al. 2005) found high (77%) heterogeneity in effect estimates from the 48 cities they selected for their meta-analysis and that 19% were negative. Goodman (Goodman 2005) comments that “stochastic uncertainty varied considerably...which produced interval estimates that reflect the uncertainty in the heterogeneity parameter...and it is not hard to imagine other situations in which a difference of that magnitude could make a qualitative difference in the inference.”

The combination of these results suggest that despite the bias for positive associations, a substantial fraction of the studies are negative and there is a lack of homogeneity. The direct test on consistency of results (different authors get similar results) fails to meet the test as different authors get different results. The evidence is suggestive of a lack of consistency, which adds to the weight of evidence against a causal association.

Table 1
Illustration of Heterogeneity of Results Within a Study, and Study to Study Variability

City	City Specific Estimate of Percent Change in Total or CVDRESP Mortality per 10-ppb change in ozone levels, with (95% CI)	
	CVDRESP Mortality (Huang, et al, 2005)	Total Mortality (Bell, et al, 2004)
Santa Ana/Anaheim	-3.03 (-6.40, 0.35)	0.31 (-0.25, 0.63)
San Bernardino	-0.98 (-3.87, 1.91)	0.61 (-0.34, 1.58)
Chicago	-0.39 (-2.38, 1.69)	0.91 (0.20, 1.57)
Pittsburgh	-0.05 (-3.17, 3.07)	0.77 (-0.19, 1.67)
Phoenix	0.05 (-3.71, 3.81)	0.83 (-0.20, 1.78)
San Antonio	0.22 (-3.84, 4.28)	0.01 (-0.84, 1.11)
Miami	0.68 (-3.51, 4.86)	0.37 (-0.19, 0.56)
Atlanta	0.77 (-3.45, 4.99)	0.20 (-0.72, 1.34)
Los Angeles	0.79 (-0.69, 2.28)	0.20 (-0.30, 0.87)
Houston	1.11 (-1.62, 3.84)	0.83 (0.02, 1.55)
San Diego	1.58 (-1.50, 4.67)	0.12 (-0.65, 1.02)
Detroit	1.88 (-1.52, 5.27)	0.77 (-0.22, 1.75)
New York	2.33 (0.93, 3.73)	1.71 (1.09, 2.31)
Dallas/Fort Worth	2.35 (0.05, 4.66)	1.10 (0.40, 1.76)
Philadelphia	2.42 (-0.22, 5.07)	1.27 (0.34, 2.12)
Cleveland	2.49 (-0.62, 5.60)	0.61 (-0.37, 1.66)
Oakland	7.97 (0.58, 15.36)	0.53 (-0.54, 1.79)
Seattle	8.08 (2.94, 13.21)	0.45 (-0.62, 1.61)
San Jose	8.18 (0.61, 15.75)	0.58 (-0.50, 1.67)

Negative values imply ozone exposure decreased acute mortality risk

3.4.4 Temporality

Temporality is defined by EPA as the occurrence of a cause before its purported effect, and is most relevant to studies of diseases that develop over a time.

EPA Staff Paper Position

- The Staff Paper suggests that for the acute health outcome of cardiovascular mortality associations reported contemporaneously (same day or 1 day lag) are expected.

ExxonMobil Comments on the EPA Staff Paper Position

The EPA position fails to note that for some studies, use of 0 day lags might mean that the effect (cardiovascular mortality) may have preceded the air pollution.

3.4.5 Lag Structure in Short-Term Studies

EPA Staff Paper Position

- Results from single day lags or distributed lags are appropriate to use for risk assessment.

ExxonMobil Comments on the EPA Staff Paper Position

While EPA indicates that 0 and 1 day lags are most plausible, they use distributed lags out to 7 days in their assessment. EPA must clearly separate single day lag results from those from distributed lag models. As clearly indicated by CASAC member Lianne Sheppard, these two models estimate separate things.

The approach used by EPA borders on lag-mining. They use results from distributed lag models since they sometimes report higher effect estimates. However, they provide no biological justification for this approach. They also use results from single day lags, again with scientific justification.

Finally EPA indicates that use of single day lags underestimated effects. As pointed out by Goodman (2005), since there is no *a priori* biologic model for estimating a plausible lag or latency, the approach generally taken is to pick the largest or strongest effect estimate. As a result, the single-point estimate derived from this approach is biased upward, not downward as suggested by EPA. As Ito et al (Ito et al. 2005) noted, “the combined estimates...may be biased upward because the ‘optimal’ or ‘best’ lags were chosen.” But the ‘best’ estimate is not based on biology nor on the ‘best’ model if that can be defined, but on a positive association.

The Agency has failed to identify biologically based lags that consider the mechanistic considerations of ozone exposure. As indicated by CASAC members John Baumes and Henry Gong, EPA has vastly over-interpreted and mis-interpreted the findings from recent studies that evaluated the cardiovascular effects of ozone. Therefore, all of the discussion concerning the timing of cardiovascular effects should be removed from this section.

3.4.6 Concentration Response Relationships and Potential Thresholds

EPA Staff Paper Position

There is insufficient evidence to support use of a threshold in quantitative risk assessment.

ExxonMobil Comments on the EPA Staff Paper Position

As noted in the above comments, ambient ozone measurements do not provide a valid surrogate for aggregate personal ozone exposure. For this reason, Brauer et al. (2002) reported that for pollutants such as ozone exhibiting low correlation between ambient and personal exposure, *it is not possible* to determine thresholds in ecological air pollution studies. Thus, it is not possible for EPA to reach the conclusion that threshold's have not been observed in ozone ecologic studies. EPA is confusing the concept "it is not possible to determine thresholds" with the concept "threshold have not been observed."

EPA has mis-interpreted the results by Vedal et al. (2003). While these authors report a statistically significant findings at 23 ppb ozone, they raise the concern that the association may be due to other factors in the air pollution mix such as particulates. Additionally, they raised the concern that measurement error attributed to use of a few ambient monitors may blur the existence of a threshold. They also noted that air pollutants may be a surrogate for meteorological factors. None of these factors are mentioned in the EPA summary of this paper.

The time series literature provides evidence that the mortality exposure response function is non linear and exhibits a threshold, as expected from extensive exposure response data from human clinical and toxicology studies. In most of the time series literature, much stronger associations in summer are observed when ozone levels are higher, than in winter when ozone is lower; in most cities there was no association in the winter. For example, Levy reported an estimate of 0.43% increase in mortality per 10 ug/m³ increase in 1-hour ozone versus -0.02% in winter. Ambient ozone concentrations in winter are lower and with a narrower range than in summer. If ozone effects are "independent" of temperature as suggested by Bates {Bates, 2005}, then the lack of an association at the lower winter concentrations is suggestive of a threshold.

Non linear exposure responses, with much higher effects at higher concentrations, and no effects at lower concentrations, are observed in human clinical and toxicology studies. For this reason, EPA has identified thresholds below which minor pulmonary function changes will not be extrapolated. It is inconsistent for EPA to not extrapolate minor pulmonary function changes below 50 ppb, but to extrapolate mortality, a much more serious health endpoint presumably related to a pulmonary mechanism of toxic action, below 50 ppb and down to "policy relevant background" (PRB) levels. EPA must use the extensive human clinical and mechanistic data to derive biologically based cutoffs for extrapolating acute mortality. In our view, these values fall in the range of 60-80 ppb.

Susceptible subgroups of the population including asthmatics, persons with reactive airway disease, cardiovascular disease, and obstructive airway disease, have been exposed to ozone concentrations of 80 ppb or higher with no or only transient changes in lung function and little or no indication of cardiac effects. Ozone is a pulmonary irritant and exposures produce irritant effects at much lower concentrations than those which could produce death. The lack of less severe and transient effects at concentrations 80 ppb is a strong argument against the hypothesis that current ambient concentrations, down to policy relevant background levels, are causing death.

As noted by EPA, Kim et al. (2004) report thresholds for acute ozone mortality in a time-series study of Seoul Korea. There are a number of nuances in the findings by Kim et al. (2004). First, higher effect estimates were reported in winter versus summer. This finding is opposite that normally observed. Second, Kim et al. did not consider fine PM as a potential confounder. Rather, their analysis was based on consideration of only coarse PM which is known to have a much lower correlation with ozone. Therefore, much of the risk attributed to ozone, either above or below the “threshold”, could have been due to confounding with fine PM. Finally their conclusion that linear model may underestimate risks does not consider actual population exposure, i.e., the number of people exposed under the threshold. Rather, this conclusion is based only a small increase in relative risk observed above the threshold.

3.5 Biological Plausibility and Coherence of Evidence

EPA Staff Paper Position

It is unclear from this section what the Agencies position is in regard to biological plausibility. This section of the Agency provides a generic table with various biological and physiological effects produced in animal and humans. There is no indication at what exposure levels these findings are evident, or how they may relate to the acute mortality with no threshold hypothesis. However, the Agency infers that coherence has been established based on morbidity findings and toxicology data.

ExxonMobil Comments on the EPA Staff Paper Position

The Agency fails to note the clear conclusion in the CD that there is no biological mechanism to explain how low levels of ozone are expected to result in acute mortality, even at policy relevant background levels. EPA should clearly note this lack of mechanistic support. Further as clearly noted by CASAC members Balmes and Gong, EPA has over-stated the evidence on cardiac effects of ozone. Again, EPA should clearly state that the available evidence does not indicate cardiac-related mechanisms of action for ozone. Overall, EPA should clearly state that there is a lack of biological plausibility for the acute ozone mortality hypothesis, particularly at low levels.

While ozone is a respiratory toxicant, the associations for respiratory mortality are *lower* than those for cardiopulmonary mortality and generally *not statistically significant*. From a coherence perspective, this is counter-intuitive.

Exposure to ambient particular matter is assumed by EPA to cause cardiopulmonary mortality. However, according to EPA, exposure to particulate matter does not impact the signal for cardiopulmonary mortality hypothesized to be caused by ozone. Therefore, one must assume that the cardiopulmonary mortality alleged to be caused by ozone occurs through a completely different biological mechanism of action than the cardiopulmonary mortality alleged to be caused by PM. However, EPA asserts both ozone and PM cardiopulmonary mortality is mediated by the same same cardiac measures of effect such as “reduced heart rate variability”.

While one would surmise that the very elderly would be most affected, in fact, the mortality association for the very elderly (>75 years) was not increased relative to the general population, or relative to younger age groups (<65 years). Rather, inexplicably, the 65-74 old age group had the highest mortality association.

Further, while the morbidity risks such as pulmonary function changes are based on solid scientific evidence, the morbidity findings from ecological epidemiology studies (hospital and emergency rooms visits) are much less certain. Further, the strength of the ecological morbidity associations is about the same as those for mortality, arguing against coherence.

We suggest that the Agency present a more scientifically balanced interpretation of entire evidence using the Hill criteria, as presented below in Table 2. A more balanced interpretation would lead to the overall conclusion that causality has not been established.

**Table 2
Overall Evidence for a Causal Relationship for Acute Ozone Mortality**

Hill Criteria	Evidence Assessment	Strength of Evidence	Comments
Strength of Association	Very Weak	Inadequate	RR 1.00-1.08 versus 1.5-2.0
Robustness	Very high	Inadequate	Ambient measurements are a very poor surrogate for aggregate personal exposure.
-exposure error	Very likely	Inadequate	Fine PM highly correlated with ozone and alleged to cause the same effects. Coarse PM controlled in most studies.
-confounding by pollutants	Moderate-High	Inadequate	Existing weather specification inadequate and produces 2-fold change in risks. Seasonality not adequately specified. Air conditioning is significant risk modifier
-model specification concerns			
Consistency	Low	Inadequate	High city to city heterogeneity within multi-city studies. High author to author heterogeneity in single city results. Cities with lowest ozone levels tend to have highest risks per unit ozone and vice versa.
Temporality	Sufficient?	Limited	Strongest signal at lag 0 is biologically plausible but may indicate effect proceeds exposure
Lag Structure	Variable, not defined	Inadequate	Multiple single day and distributed lags used interchangeably. Biologically based lags not established.
Biological Plausibility	Not established	Inadequate	Not supported by very extensive human clinical database. No biological mechanism to explain acute mortality. Plausibility especially low at low ozone concentrations.
Coherence	Low	Limited	Pulmonary function effects derived from solid scientific data, but are transient and not expected to result in mortality. Morbidity risks from ecological studies (hospital admissions, emergency room visits) are less certain and strength of association's in some cases are about the same as per mortality. Respiratory mortality is not statistically significant. Mortality risks are not higher for the very elderly. PM does not modify ozone cardiopulmonary mortality but alleged to be mediated by similar cardiac effects (e.g., decreased heart rate variability).
Overall		Inadequate	Causality clearly not established

Section 5.3.2.3. Concentration-Response Functions

EPA proposes to use concentration response functions (CFRs) from individual cities based on the logic that “a CFR estimated in the assessment location is preferable since it avoids uncertainties related to potential differences due to geographic location.” This rationale is flawed, since the uncertainty introduced by the statistical model used by an individual author of a single time series city study is much greater than the uncertainty introduced by location specific factors. Further, the logic contradicts EPA’s own conclusion, namely, that the acute ozone mortality estimates are “robust” to metrological factors, to confounding by co-pollutants, and to other factors that differ by location. In reality, the results of single city time-series studies are notoriously heterogeneous and model dependent.

Marked variability in results for single studies have also been observed in multi-city studies and in meta analyses. This can be illustrated by the marked variation in results reported in some of the “key” studies identified by EPA (see figure 1). For example, for the city of Santa Ana/Anaheim, would EPA suggest using the results presented in Bell et al. (0.4% *increased* risk per 10 ppb ozone), or those from Huang et al. (3.03 *decreased* risk per 10 ppb ozone)? Similar variability is observed for many other cities. This variability renders the results of single studies, either as reported in individual publication or in multi-city studies, as useless for purposes of risk assessment.

The three meta-analyses indicated the possibility of a small increased risk of mortality with increasing ozone exposure. However, a close examination of the caveats and concerns provided by the authors and reviewers of these studies reveals that the principle contribution of the three meta analyses were to highlight the existence of publication and selection bias, heterogeneity, and model dependence bias in the existing ozone time-series study results. One cannot conclude from these studies alone that minor variations in ambient ozone *causes* acute mortality. Further, given all the uncertainties, the risk values presented in the meta analyses are clearly over-estimates and should not be used to quantify acute ozone mortality in future regulatory-based ozone cost benefit evaluations. We present below some of the concerns noted by the authors of these studies, along with some additional concerns we have identified.

Author and Reviewer Concerns

- Levy noted that the meta-analysis has limited interpretability because of the different analytical methods used in single-city studies. As a result, he suggested that the pooled estimates from the meta analysis were hard to interpret or apply in a regulatory impact analysis.
- All of the author’s noted the high degree of heterogeneity, or variability, in the estimates from individual cities used in the 3 meta analyses, and the fact that the results for many of the individual cities did not reach statistical significance. For an illustration of these problems, see table 1. Since the meta-analyses approach assumes homogeneity, the high variability resulted in an underestimate of the confidence interval, or uncertainty, in the risk estimates. This error in turn led to an overestimate of the statistical significance of the estimated risks.
- Goodman noted that the risk estimates from single-city studies used in the meta-analysis were biased ~3-fold upwards because the strongest signal was used.

- Bell noted that publication bias was apparent with the meta analysis approach, which resulted in a marked increase in the risk estimates versus those from multi-city studies. This bias results from the preference for publishers to accept for publication studies reporting health effects, while rejecting studies not reporting effects. When the results were corrected for publication bias, the estimated risk was reduced from 0.39%(0.26-0.51%) to 0.35% (0.23-0.47%) per 10-ppb increase in 1-hour ozone, or about 10% (Ito).
- The authors and reviewers observed that process of estimating the risks was complex, since different modeling methods were used. Further, the size of the risk estimate was influenced by the methods used. Some items that varied across the 3 analyses included:
 - Lag times, or the time between the rise in pollution and the health effect
 - Ozone metric considered (1 hr max, 8-hr mean, 24-hr mean, etc)
 - Rules to convert one metric to another (Levy used 4:3:2, Bell used 20:15:8)
 - Age of decedent
 - Mortality type (total, cardiovascular, pulmonary)
 - Adjustment for weather, season, air conditioning, other pollutants

Interestingly, Levy noted that if studies that did not adequately address all potentially influential factors were excluded, almost no studies would have been available for the meta analysis. Ito noted that alternate weather models produced a 2-fold difference in risk estimates. Bates noted that the correlation between temperature and ozone makes it difficult to separate effects of these two factors. Levy noted that further work is needed to explore potential confounding of the ozone mortality relationship by PM_{2.5}, which is highly correlated with ozone in summer and is also associated with acute mortality, and air conditioning. Failure to adequately adjust for this would result in double counting of acute mortality from ozone and PM_{2.5}.

- The overall estimate of ~0.8% increased mortality per 10 ppb increase in average daily ozone suggested by Bates upon review of the meta analyses does not adjust for other effects and bias, and so is an overestimate of the overall observed associations. Adjustment for some of these factors produces different estimates of effect.
 - An overall estimate of ~0.25% increase in mortality/10 ppb of 24-hour average ozone is the more appropriate estimate from Bell et al.
 - Adjustments in Ito et al {Ito, 2005} suggest the overall estimate is about 0.5% or less. The 4-smoother weather model used in other major time series studies such as the National Mortality, Morbidity Air Pollution Study (NMMAPS) resulted in a 0.5% risk estimate using the average of 0- and 1-day lags. The more aggressive adjustments for weather seem appropriate given that they tend to give better fits to the data. The overall estimate is about 0.3% if the model also contains PM. Levy et al (2005) indicated that selection of a 1-2 day lag instead of a 0-day lag resulted in an overall estimate of ~0.4% increase in mortality per 10 ppb increase in 24-hour O₃. Further adjustments such as for weather, publication, PM or NO₂ could further reduce the effect estimate.

Section 5.3.2.5 Characterizing Uncertainty and Variability

EPA claims that they have addressed uncertainty in the CFRs, since statistical uncertainty surrounding the CRFs are reflected in the confidence intervals. Statistical uncertainty considers whether the findings are due to chance (p value or confidence intervals), and is the only source of uncertainty that is considered by EPA. However, it is not the most important source of uncertainty and is largely dependent on sample size. The increase in sample size brought about by combining many studies in the meta-analyses so increased statistical power that even

the minuscule effect estimates became statistically significant and were allowed to overshadow the more important hidden uncertainties. For example, after excluding extreme values, Levy reported that only 18 of 46 reported values for individual cities were statistically significant, yet all 46 were included in the meta-analysis. Hidden uncertainties are a more important source of uncertainty and have a significant impact on the effect estimates, generally to further reduce the strength of the associations.

EPA indicates that there is uncertainty in whether or not a causal relationship has been established for acute ozone mortality. However, EPA does not consider this uncertainty *at all* in the risk assessment. Through this approach, EPA assumes 100% probability that the associations are causal. As summarized above, and as indicated by CASAC, the causal relationship is far from conclusive. To capture the full range of uncertainties, EPA should present a risk assessment based on the possibility that causality has not been established for acute ozone mortality.

EPA indicates there is uncertainty in if the shape of the CRF for acute ozone is linear at low concentrations with no threshold. Again, even though EPA states that the evidence for the linear no threshold approach is based on "limited" data, EPA does not consider this uncertainty *at al*. To capture the full range of uncertainties, we suggest evaluating cutpoints of 80, 70, and 60 ppb below which acute ozone mortality cannot be quantified.

References

Bates, D.V. (2005). Ambient ozone and mortality. *Epidemiology* 16, 427-429.

Bell, M.L et al. (2005). A meta analysis of time-series studies of ozone and mortality with comparison to the National Morbidity, Mortality Air Pollution Study. *Epidemiology* 16, 436-445.

Brauer et al. 2002). Exposure misclassification and threshold concentrations in a time series analyses of air pollution health effects. *Risk Analysis* 22 (6), 1183-1193.

Goodman, S.N (2005). The methodological ozone effect. *Epidemiology* 16, 430-435.

Hill, A.. (1965). The environment and disease: association or causation? *Proc. Royal Society Medicine* 58, 295-300.

Huang et al. (2005). Bayesian hierarchical distributed lag models for summer ozone exposure and cardio-respiratory mortality. *Environment* 16, 547-562.

Ito, K et al. (2005). Associations between ozone and daily mortality: analysis and meta – analysis. *Epidemiology* 16, 458-468.

Kim et al. (2004). Determining the threshold effect of ozone on daily mortality: an Analysis of ozone mortality in Seoul, Korea. *Environ. Research* 94, 113-119.

Levy et al. (2005) Ozone exposure and mortality: an empiric Bayes Metaregression Analysis. *Epidemiology* 16, 458-468.

Phillips, C.V et al. (2004). The missed lessons of Sir Austin Bradford Hill. *Epidemiologic Perspective and Innovations* 1, 1-19.

Vedal et al. (2003). Air pollution and daily mortality in a city with low levels of pollution. *Environ Health Perspect* 111, 45-51.