Comments on Second Draft Ozone Integrated Science Assessment

Docket ID No. EPA-HQ-ORD-2011-0050

Julie E. Goodman, Ph.D., DABT

Sponsored by
American Petroleum Institute
1220 L Street NW
Washington, DC 20005-4070

December 29, 2011

Gradient

Table of Contents

Page

Executive Summary ES-1

1 Introduction 1

2 Assessment of the Weight of Evidence 2
- 2.1 The ISA Does Not Appropriately Evaluate the Weight of the Evidence for Health Effects of Ozone 2
- 2.2 The ISA Does Not Fully Consider All of the Bradford Hill "Aspects" in its Causal Determinations 4
- 2.3 The ISA Does Not Weigh Alternative Views 5
- 2.4 Many of the Same Issues Identified in Other EPA Assessments Are Apparent in the Ozone ISA 5
- 2.5 EPA Should Follow a Systematic and Rigorous Approach for Assessing the Evidence 8

3 Assessment of Controlled Exposure Studies 10
- 3.1 A Cross-Study Comparison and Mode-of-Action Data Indicate a Threshold for Effects of Ozone on Lung Function 10
- 3.2 Mean FEV_1 Decrements at 60 ppb Are Not Statistically Significant in All But One Study and Do Not Support Causation 13
- 3.3 EPA Does Not Fully Consider Adversity Criteria 14
- 3.4 Controlled Human Exposure Studies Were Not Designed to Assess Individual Data 17

4 Conclusions 20

References 21

Appendix A Data for Evaluation of an Ozone Concentration Threshold for Pulmonary Function Effects After Acute Exposure in Healthy Subjects

Appendix B Gradient's Comments on First Draft Ozone Integrated Science Assessment

Executive Summary

In September, 2011, the United States Environmental Protection Agency (EPA) released the second external review draft of the *Integrated Science Assessment for Ozone and Related Photochemical Oxidants* (referred to here as the "Second Draft ISA"). Gradient previously identified several issues with EPA's approach for evaluating controlled human exposure, short-term epidemiology, and long-term epidemiology studies of ozone and the causal determinations based on these evaluations in the First Draft ISA for ozone. These issues are still evident in the Second Draft ISA.

Although EPA claims to use a weight-of-evidence approach, with consideration of the Bradford Hill "aspects," there is no evidence that such an approach was used as no weight is assigned to the studies. The biological gradient, or exposure-response relationship, is often overlooked in EPA's assessment of the epidemiology studies, as is the aspect of the strength of association. EPA does not consider evidence supporting the alternative view that it is not ozone, but another factor, that caused the health effects associated with ozone in some of the epidemiology studies. EPA also does not consider other scientifically accepted statistical methods besides those used in EPA's reanalysis of the controlled human exposure data.

Many of these issues have been identified in other EPA assessments by a National Research Council (NRC) committee, such as the lack of methods for evaluating the strengths and limitations of studies, leading to evaluations that appear to give equal weight to all publications. The implementation of weight-of-evidence guidelines appears to be subjective, and variable detail is provided for how the criteria are applied. There also appears to be no evaluation of the consistency of findings within and among studies, and positive associations are overemphasized, whether they are statistically significant or not. EPA's causality criteria are not applied in a thorough and consistent manner throughout the assessment. EPA should follow the NRC best practices recommendations in a revision of its weight-of-evidence approach.

The hypothesis-based weight-of-evidence (HBWoE) approach is an example of a rigorous approach for assessing evidence that is consistent with both EPA risk assessment guidance documents and the NRC committee's best practices and has been successfully applied in evaluations of potential health effects for several chemicals. The HBWoE approach provides a structure for the process of evaluating the judgments that must be made in a weight-of-evidence evaluation, and is a means to

rigorously and transparently express judgments about whether and to what extent the Bradford Hill aspects are met. EPA should consider such an approach in its assessments of ozone and other criteria pollutants.

In the First and Second Draft ISA, EPA assessed controlled exposure studies of ozone in healthy adults. Contrary to EPA's assessment, these studies indicate a threshold below 70 parts per billion (ppb) at which there are no statistically significant adverse effects associated with ozone, and this is consistent with biological data that support a threshold mechanism of action. Effects at 60 ppb are not adverse, nor do they occur more often than do those associated with filtered air (FA) exposures. In addition, any large decrements in lung function in particular individuals at this exposure level cannot be attributed to ozone.

Overall, the Second Draft ISA does not follow an objective and rigorous weight-of evidence approach for evaluating the available data and does not evaluate the controlled human exposure data in a scientifically appropriate manner. Because of this, the ISA cannot be relied upon to support a causal relationship between ozone exposure and health outcomes at exposure levels below the current National Ambient Air Quality Standards (NAAQS). EPA should follow the NRC recommendations for best practices in risk assessment to ensure its goal of providing credible causal conclusions about ozone-related health effects in the final ISA.

1 Introduction

In September 2011, the United States Environmental Protection Agency (EPA) released the second external review draft of the *Integrated Science Assessment for Ozone and Related Photochemical Oxidants* (referred to as "Second Draft ISA" throughout these comments) (US EPA, 2011a). Gradient provided comments (Goodman, 2011; provided in Appendix B) on the First Draft ISA, released in March 2011 (US EPA, 2011b), and little has changed in the Second Draft ISA. There are still several issues associated with EPA's approach for evaluating controlled human exposure, short-term epidemiology, and long-term epidemiology studies of ozone and the causal determinations that were made in the ISA for the relationship between ozone exposure and the health effects examined in these studies.

Below, we evaluate EPA's assessment of the evidence for ozone-related health effects as presented in the Second Draft ISA. Overall, this assessment does not follow an objective and rigorous weight-of evidence approach for evaluating the available data and does not evaluate the controlled human exposure data in a scientifically appropriate manner. Because of this, the ISA cannot be relied upon to support a causal relationship between ozone exposure and health outcomes at exposure levels below the current National Ambient Air Quality Standards (NAAQS).

2 Assessment of the Weight of Evidence

The process for developing an ISA should involve an evaluation of the weight of the evidence, as stated in the Preamble of the Second Draft ISA (US EPA, 2011a):

> EPA integrates the evidence from across scientific disciplines or study types and characterizes the weight of evidence for relationships between the pollutant and various outcomes (p. lii).

In addition, EPA's framework for causal determination in the ISA is based on modified Bradford Hill "aspects," and EPA (2011a) states:

> [O]ne cannot simply count the number of studies reporting statistically significant results or statistically nonsignificant results and reach credible conclusions about the relative weight of the evidence and the likelihood of causality. Rather, these aspects are taken into account with the goal of producing an objective appraisal of the evidence, informed by peer and public comment and advice, which includes weighing alternative views on controversial issues (p. lxi).

EPA did not follow through with this in either the First Draft or Second Draft ISA, however. As discussed in more detail below, there are no clear statements indicating how the evidence is weighed, how evaluations of the various Bradford Hill aspects are to be carried out, how judgments regarding whether aspects are to be considered as met are made, and how an overall judgment is to be synthesized out of consideration of all of the aspects jointly. EPA also does not follow through with weighing alternative views on controversial issues, and does not fully consider whether reported effects are adverse. Thus, the ISA framework does not provide an objective appraisal of the evidence.

2.1 The ISA Does Not Appropriately Evaluate the Weight of the Evidence for Health Effects of Ozone

In Gradient's comments on the First Draft ISA (Goodman, 2011), it was noted that there is no consistent discussion of strengths and limitations of individual epidemiology studies and, thus, no "weight" is assigned to the studies. Rather, EPA provides summarizing statements that emphasize only the few positive associations for a given outcome. Many examples of this are still evident in the Second Draft ISA, as discussed below.

In the discussion of respiratory effects in adult day-hikers in Section 6.2.1.2, the positive associations with lung function decrements reported in the study by Korrick *et al*. (1998) are emphasized over the null associations for the same endpoints reported in the study by Girardot *et al*. (2006), with no discussion of the strengths and limitations of each study, with the exception of noting the speculation by Girardot *et al*. (2006) as to why their results may have differed from those of Korrick *et al*. (1998). EPA should provide its own analyses of the strengths and limitations of the epidemiology studies and use these analyses to assign weight to the various studies evaluated in the ISA.

In its evaluation of studies examining short-term ozone exposure and cause-specific mortality in Section 6.6.2.5, EPA states that there is evidence of associations with cardiovascular- (CV) and respiratory-specific mortality, yet all risk estimates for respiratory mortality and almost all for CV mortality were null in all-year analyses. In addition, summer analyses reported mixed results for both respiratory and CV mortality. EPA does not provide any discussion of the strengths and weaknesses of these studies, or how null results were considered in the weight-of-evidence analysis, to support its conclusion that they provide evidence of an association.

In its summary of epidemiology data for short-term effects on pulmonary inflammation and oxidative stress in Section 6.2.3.2, EPA states that many recent studies reported positive associations, yet throughout that section, EPA states that the results are very mixed and inconsistent, and does not show that studies with positive associations carry more weight than those reporting null associations. EPA should not emphasize the positive associations over the null ones.

In the discussion of epidemiology studies of short-term effects on respiratory symptoms in Section 6.2.4, EPA states there is a "strong" body of evidence demonstrating associations in asthmatic children and that there is evidence of an association in asthmatic adults as well, yet almost all risk estimates are null for this outcome in both single- and multi-pollutant models, and there is no discussion of study strengths and weaknesses to support the overemphasis of the few studies with positive results.

As noted in Gradient's comments on the First Draft ISA (Goodman, 2011), EPA's evaluation of the epidemiology data often discounts or misinterprets evidence from studies with null results, and this is still evident in the Second Draft ISA. For example, in Section 6.2.4.1, EPA discounts the null results of the study of respiratory symptoms in children with asthma by Schildcrout *et al*. (2006), noting that the

study may have lacked sufficient power to detect ozone-related effects because the authors restricted their analyses to the summer months, but relies on the study by Mortimer *et al*. (2002), which was conducted only during the warm season. In Section 6.2.1.2, an example of how EPA misinterprets null evidence by classifying studies with positive, but very small and not statistically significant, risk estimates as evidence for a causal relationship can be seen. EPA claims that the results of the study by O'Connor *et al.* (2008) support its conclusion that current levels of ozone cause lung function decrements in asthmatic children. EPA states that using year-round data, O'Connor *et al*. (2008) reported decreases of 0.41% and 0.22% in forced expiratory volume in 1 second (FEV_1) and peak expiratory flow (PEF), respectively, per 20 parts per billion (ppb) increase in 24-hour average ozone concentrations (for the average of lags 1-5). EPA does not note in the ISA that neither of these results were statistically significant or that the magnitude of the response for ozone was small compared to that reported for other pollutants. EPA also fails to note that O'Connor *et al*. (2008) reported no association between ozone and asthma symptoms or school absences, although in Section 6.2.4.1, the reported null association with asthma symptoms (wheeze) is incorrectly stated as being positive. Although a group of studies with positive risk estimates for a given endpoint that did not achieve statistical significance could still indicate a true association, EPA did not do a meta-analysis to confirm this for any endpoint. Without a meta-analysis, results of such studies should only be considered null, and do not provide any evidence of an association.

In summary, EPA does not appropriately evaluate the weight of the evidence for ozone-related health effects in the Second Draft ISA, despite its intent to do so. As in the First Draft ISA, positive results are overemphasized without assigning weight to studies or even discussing the strengths and limitations of each study evaluated, and null results are often discounted or used as positive evidence of a causal relationship without a meta-analysis to confirm whether such an interpretation is correct.

2.2 The ISA Does Not Fully Consider All of the Bradford Hill "Aspects" in its Causal Determinations

Both the First and Second Draft ISA do not fully consider all of the Bradford Hill "aspects" in the evaluation of the potential health effects of ozone. For example, the aspect of a biological gradient, or exposure-response relationship, is often overlooked in EPA's assessment of the epidemiology studies. In Section 6.2.1.2 of the Second Draft ISA, EPA notes that there is no exposure-response relationship for lung function changes across short-term exposure studies of outdoor workers, indicating a lack of causality. EPA does not mention this deviation from the Hill aspects in its summary and causal

conclusion for short-term respiratory effects, however. The ISA also does not fully consider the Hill aspect of the strength of association, as EPA notes that the lung function changes were small in the outdoor worker studies, but this does not factor into the overall judgment of causality for respiratory effects of ozone.

2.3 The ISA Does Not Weigh Alternative Views

The Preamble of the Second Draft ISA states that EPA will weigh alternative views on controversial issues in the evaluation of ozone-related health effects; however, EPA does not follow through with this in the ISA. The Preamble notes that consideration of the limitations of epidemiology studies, such as potential confounding, effect modification, and exposure measurement error, must be taken into account to properly inform their interpretation, yet EPA does not consider these limitations in its "weight-of-evidence" conclusions for causality. If EPA were truly to weigh alternative views, it would consider the view that it is not ozone, but another factor, that is causal for the health effects associated with ozone in some of the epidemiology studies. It would then discuss the reasons why this view is less likely to be true than the view that ozone is the causal factor. In addition, as described in more detail in the controlled exposure section below, EPA's evaluation of controlled exposure studies with exposures to 60 ppb ozone in Section 6.2.1.1 of the Second Draft ISA does not consider other scientifically accepted statistical methods besides those used in EPA's reanalysis of the data.

2.4 Many of the Same Issues Identified in Other EPA Assessments Are Apparent in the Ozone ISA

As noted in Gradient's comments on the First Draft ISA (Goodman, 2011), many of the issues associated with EPA's assessment of the evidence for ozone-related health effects in the First Draft ISA have been identified in other EPA assessments, and these issues still remain in the Second Draft ISA. For example, a committee assembled by the National Research Council (NRC) of the National Academy of Sciences (NAS) recently published its *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde* (NRC, 2011), noting a number of deficiencies that precluded confidence in the weight-of-evidence conclusions drawn in the Integrated Risk Information System (IRIS) assessment, as well as general issues that other NRC committees have identified in a number of EPA's IRIS assessments over the last decade. The committee stated that the EPA draft formaldehyde assessment

"was not prepared in a transparent, consistent fashion," and the committee criticized EPA's hazard identification assessment, noting many issues related to EPA's approach for weighing the evidence. These issues included an inconsistent method for evaluating strengths and weaknesses of studies and the lack of a clear framework for evaluating the weight of the evidence in establishing causation, which are also apparent in EPA's assessment of the ozone epidemiology data in the First and Second Draft ISA.

The NRC committee noted that the methods for evaluating the strengths and limitations of the selected studies were not provided in the draft formaldehyde assessment. It also felt that the descriptions and evaluation of the individual epidemiology studies were not consistent among studies, stating, "The characterization of the strengths and weaknesses of the studies varies; some studies receive a fuller treatment, including a more extensive assessment of bias and its consequences for estimating effect measures, and others receive less attention." This inconsistency is further noted by the committee in its statement that "In some cases, there is a tendency to describe the studies ultimately selected for the derivation of the RfC [reference concentration] in favorable terms," without sufficient consideration of their weaknesses. The committee felt that although some attention was given to methodological concerns, particularly in studies considered to be more informative, it was given in a nonsystematic fashion, as some key methodological limitations were inconsistently mentioned and insufficiently explored. The committee stated that the lack of evidence for a specified format for evaluating studies led to weight-of-evidence evaluations that appeared to give equal weight to all publications, with no consideration of study quality or validity of exposure concentration measurements. As described above, this lack of consistent discussion of study strengths and weaknesses is also evident throughout the assessment of epidemiology data in the First and Second Draft ISA for ozone.

Further, the NRC committee noted inconsistencies in the approach that EPA used for evaluating causation for different health endpoints in the draft formaldehyde assessment. The committee stated that the implementation of weight-of-evidence guidelines in the assessment "appears to be subjective and not standardized." The committee found that variable detail was provided in how the weight-of-evidence criteria had been applied, noting that uniformly developed discussions applying the criteria could not be identified at appropriate points in the text. The committee also felt that EPA took an *ad hoc* approach to hazard identification for some outcomes. These same issues in the evaluation of causation are also found in the First and Second Draft ISA for ozone. Although the Preamble of the Second Draft ISA sets out EPA's causality criteria, they are not applied in a thorough and consistent manner throughout the evaluation of ozone-related health effects.

Finally, the NRC committee found that the concept of consistency of findings within and among studies was not defined in EPA's draft assessment of formaldehyde, and the committee stated that in some cases "the conclusion of causation appears to be based on a subjective view of the overall data." The committee also noted that some concluding statements regarding causation seemed to over interpret the results. The committee recommended that clear narratives are needed to provide the rationale for conclusions, and that when concluding statements are made that indicate the results support an association in the face of mixed results (*e.g.*, positive, weak, and null studies for a given endpoint), these need to be accompanied by a thorough discussion of the strengths and weaknesses of the studies. As discussed above, these same issues are found in the First and Second Draft ISA for ozone, in which positive results (regardless of their statistical significance) are overemphasized without a discussion of the strengths and limitations of studies and without assigning weight to the studies.

Overall, the NRC committee noted recurring methodological issues in the draft assessment of formaldehyde and provided many recommendations for revision. The committee stated that a clearly articulated framework for weighing the evidence is critical for any determination of causation. It recommended that all key studies "need to be thoroughly evaluated with standardized approaches that are clearly formulated based on the type of research." It also recommended that the approach for weight-of-evidence evaluation in the assessment be revised. The committee stated, "Strengthened, more integrative, and more transparent discussions of weight of evidence are needed. The discussions would benefit from more rigorous and systematic coverage of the various determinants of weight of evidence, such as consistency." Although the causality framework and the intended methods of assessment are clearly stated in the beginning of each ISA, the same issues noted by the NRC committee are evident in the implementation of the framework in the First and Second Draft ISA for ozone. It is noteworthy that the CASAC review panel for the First Draft ISA for lead recommended a "more rigorous and transparent 'weight of the evidence' analysis" that should "devote more attention to the limitations of the existing studies with respect to consistency, reproducibility, bias, control for confounders, and shortcomings in statistical methodology" in an October 25, 2011 draft letter to EPA (Frey and Samet, 2011), indicating that these issues are not limited to the ISA for ozone, but are evident in EPA assessments of other criteria pollutants as well. EPA should follow the recommendations of the NRC committee and CASAC in a revision of its weight-of-evidence approach to ensure its goal of providing credible causal conclusions about ozone-related health effects in the final ISA.

2.5 EPA Should Follow a Systematic and Rigorous Approach for Assessing the Evidence

As noted in the sections above, EPA did not appropriately evaluate the weight of the evidence for ozone-related health effects in the First and Second Draft ISA. EPA should consider that a weight-of-evidence analysis is more than just a matter of describing a set of available studies with an array of results and then announcing an overall professional judgment. It is important to be systematic and transparent about the information being drawn from the studies, describing the methodology used for the evaluation and formulation of judgments, and the scientific reasoning behind any judgments offered. The principles and procedures for such an analysis should be in place before it is performed so that the criteria are clear and independent of the particular set of studies at hand and to ensure that a rigorous approach for evaluating the evidence is taken. Such a systematic and rigorous approach was recently described by the NRC committee that reviewed EPA's draft IRIS assessment of formaldehyde. The committee set out a "roadmap" for best practices in risk assessment, highlighting what it considers to be the critical aspects for the development of a scientifically sound assessment. These aspects include:

1. An overall, documented, and quality-controlled process for assessment;
2. Standardized approaches for reviews of various types of studies to ensure uniformity;
3. A standardized presentation of reviewed studies in tabular or graphic form to capture key dimensions of study characteristics and weight of evidence;
4. Development of templates for the evidence tables and establishment of protocols for reviewing major types of studies;
5. Standardized approach to using weight-of-evidence guidelines and workshops on approaches to implementing such guidelines;
6. Uniform language to describe the strength of evidence for effects; and
7. Expanded and harmonized approach for characterization of uncertainty and variability.

The NRC committee stated that EPA's current process for NAAQS reviews, including the development of an ISA, is a "useful" example of *changing a process* for evidence review and risk assessment in a relatively short period of time. The committee did not state that any specific ISA prepared to date is a good example of a risk assessment or follows the appropriate methods for evaluating evidence. Based on its many issues described above, it is clear that the First and Second Draft ISAs for ozone do not follow the best practices set out by the NRC committee.

The hypothesis-based weight-of-evidence (HBWoE) approach is an example of a rigorous approach for assessing evidence that is consistent with both EPA risk assessment guidance documents and the NRC committee's best practices and has been successfully applied in evaluations of potential health effects for several chemicals (Rhomberg *et al*., 2010, 2011; Prueitt *et al*., 2011). Indeed, the recently finalized EPA guidance document regarding weight of evidence in interpretation of data from the Endocrine Disruptor Screening Program, W*eight-of-Evidence: Evaluating Results of EDSP Tier 1 Screening to Identify the Need for Tier 2 Testing* (US EPA, 2011c), cites the HBWoE evaluation of naphthalene carcinogenicity (Rhomberg *et al*., 2010) as an example of successful application of a rigorous weight-of-evidence analysis.

The HBWoE approach is hypothesis-based in the sense that it emphasizes articulation of the proposed bases for the relevance of the data to the causal question at hand, specifying the logic and reasoning. The approach weighs all of the data (*e.g.*, epidemiology, animal toxicology, mode of action), both positive and negative, in terms of quality and relevance to humans in a way that allows each data set to inform the other, and further synthesizes all of the data to determine overall plausibility for causality in humans, considering uncertainties and inconsistencies in the data sets and *ad hoc* assumptions that may be required for some of the hypotheses put forth.

In the causality framework set forth in the ISA, there is a lack of clear guidance for how the evaluations of the various Bradford Hill aspects are to be carried out, how judgments regarding whether aspects are to be considered as met are made, and especially how an overall judgment is to be synthesized out of consideration of all of the aspects jointly. This is especially important in view of the NRC (2011) "roadmap" recommendations that EPA should develop and use to promulgate policies regarding the processes and considerations to be followed as it makes its weight-of-evidence judgments. The HBWoE approach suggests just this kind of structure to the process of evaluating the judgments that must be made in a weight-of-evidence evaluation, and is a means to rigorously and transparently express judgments about whether and to what extent the Bradford Hill aspects are met. EPA should consider such an approach in its assessments of ozone and other criteria pollutants.

3 Assessment of Controlled Exposure Studies

In the First and Second Draft ISA, EPA assessed controlled exposure studies of ozone in healthy adults, focusing on four controlled exposure studies that assessed the association between ozone and lung function at exposures below 80 ppb (Adams, 2002, 2006; Schelegle *et al*., 2009; Kim *et al*., 2011). Together, these studies indicate a threshold below 70 ppb at which there are no statistically significant adverse effects associated with ozone, and this is consistent with biological data that support a threshold mechanism of action. Effects at 60 ppb are not adverse, nor do they occur statistically more often than do those associated with filtered air (FA) exposures. In addition, one cannot determine whether there would be a population shift in lung function decrements at 60 ppb ozone exposure based on the small number of individuals in the controlled exposure studies. Furthermore, any large decrements in lung function in particular individuals at this exposure level cannot be attributed to ozone. These issues are discussed in more detail below.

3.1 A Cross-Study Comparison and Mode-of-Action Data Indicate a Threshold for Effects of Ozone on Lung Function

In Figure 6-1 of the First Draft ISA, EPA presented a cross-study analysis of several studies that evaluated controlled exposures between 40 and 120 ppb (Adams, 2002, 2003, 2006; Folinsbee *et al*., 1988; Horstman *et al*., 1990; McDonnell *et al*., 1991, 2007). In this figure, EPA included a smooth curve without a threshold that is based on the study by McDonnell *et al*. (2007), which is not based on exposures < 80 ppb or any data from the more recent studies by Adams (2002, 2003, 2006), Schelegle *et al*. (2009) and Kim *et al*. (2011); rather, it relies heavily on data from Folinsbee *et al*. (1988), Horstman *et al*. (1990), and McDonnell *et al*. (1991). EPA also presents this analysis as Figure 6-1(A) in the Second Draft ISA. While EPA acknowledges that effects on FEV_1 aren't statistically significant at 40 and 60 ppb in the individual studies from the figure that examined these exposure concentrations, the error bars (representing the standard error of responses, where available) in Figure 6-1(A) do not reflect this. In the Second Draft ISA, EPA added Figure 6-1(B), which focuses on FEV_1 decrements ranging from 40 to 80 ppb in the studies by Adams (2002, 2003, 2006), Horstman *et al*. (1990), Kim *et al*. (2011), McDonnell *et al*. (1991), and Schelegle *et al*. (2009) but does not include error bars or any modeled curve. EPA should include the 95% confidence intervals (CIs) around each point in Figures 6-1(A) and 6-1(B) to more accurately reflect the variability (and lack of statistical significance) at 40 and 60 ppb exposures, and

should also include models that describe the data in these figures. Gradient performed such an analysis, as discussed below.

We compiled the same dataset that EPA used in its evaluations (Table 1, attached in Appendix A) and calculated the group mean decrease in ΔFEV_1 from available ΔFEV_1 data for a given ozone concentration and the corresponding filtered air (FA) controls and estimated, where possible, the Standard Deviations (SD) of the group mean decrease in ΔFEV_1.

The group mean data and the SDs are plotted in Figures 1 and 2 below. Rather than evaluating individual data points from each of the studies separately, we fitted two different models (a linear fit in Figure 1 and a three parameter sigmoid curve fit in Figure 2) to the group mean decrease in ΔFEV_1 from across the studies. The fitted models and their 95% confidence bands are shown in the respective figures.[1] Associated regression statistics and the model parameters' values are also shown in the tables below the respective figures.

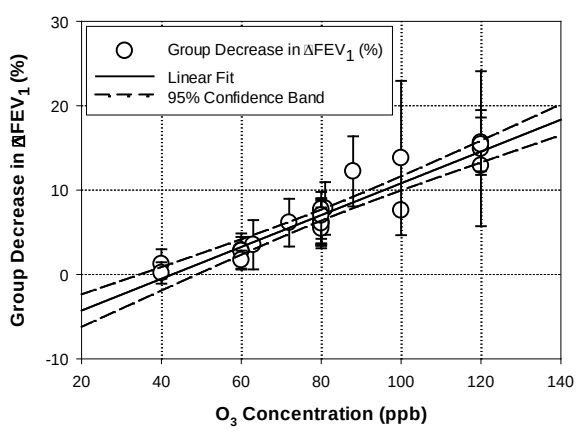

Figure 1: A linear fit to the group mean decrease in ΔFEV_1.

f = y0+a*x				
R	R^2	Adj R^2	SEE	
0.94	0.89	0.88	1.55	
	Coefficient	SE	t	P
y0	-8.05	1.20	-6.71	<0.0001
a	0.19	0.01	13.32	<0.0001

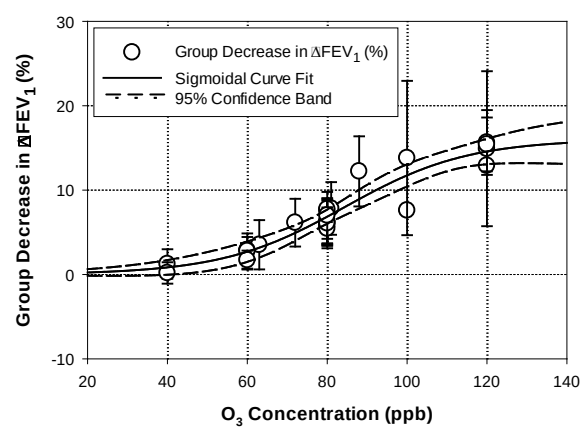

Figure 2: A three-parameter sigmoidal curve fit to the group mean decrease in ΔFEV_1.

f= a/(1+exp(-(x-x0)/b))				
R	R^2	Adj R^2	SEE	
0.95	0.90	0.89	1.50	
	Coefficient	SE	t	P
a	16.0	1.6	10.01	<0.0001
b	15.4	3.5	4.436	0.0002
x0	84.5	4.2	19.9	<0.0001

[1] The term confidence band refers to the region of uncertainties in the predicted values over a range of values for the independent variable (SigmaPlot version 12.0).

Table 2 shows the predictions of group mean decrease in ΔFEV_1 (%), at various ozone concentrations, based on the two models (Figures 1 and 2).

Table 2
Predictions of Group Mean Decreases in ΔFEV_1 (%) Based on Two Models Fitted in Figures 1 and 2

O_3 Concentration (ppb)	Predicted Group Mean Decrease in Δ FEV_1 (%)	
	Linear Model	**Sigmoid Model**
0	-8.05	0.07
20	-4.25	0.24
40	-0.45	0.84
60	**3.35**	**2.71**
80	7.15	6.84
120	14.75	14.55
140	18.55	15.58

The linear model clearly indicates an intercept at 42.4 ppb ozone—with the lower 95% confidence band closer to 50 ppb ozone (Figure 1), but a protective effect of ozone at these low concentrations is biologically implausible. The sigmoid model fits the data, indicating there is likely a threshold, and predicts a group mean decrement in ΔFEV_1 at 60 ppb ozone of 2.71%, with the lower 95% confidence band expected to be even lower (Table 2 and Figure 2). Such a small decrement is within the intra-day variability of FEV_1 in normal subjects and does not meet established criteria for a clinically adverse effect on lung function, however, as discussed in more detail in Section 3.3.

Regarding EPA's conclusion that Figure 6-1(A) indicates a smooth dose-response curve without evidence of a threshold for exposures between 40 and 120 ppb, the fact that a statistical curve without a threshold can be fit to the data does not itself provide evidence for a lack of threshold. If one has information regarding thresholds or the mode of action of an agent, this should inform the statistical curves that are fit. As discussed in Gradient's comments on the First Draft ISA (Goodman, 2011), there is evidence suggesting that antioxidant defenses against ozone indicate a threshold mode of action for effects on lung function. One should not choose a non-threshold curve when the mode of action is clearly threshold in nature.

3.2 Mean FEV_1 Decrements at 60 ppb Are Not Statistically Significant in All But One Study and Do Not Support Causation

EPA claims that the data support an association between 60 ppb ozone and lung function decrements. On page 6-11 of the Second Draft ISA, EPA (2011a) states:

> At 60 ppb, there is information available from 4 separate studies (Adams, 1998) (Kim *et al.*, 2011; Schelegle *et al.*, 2009; Adams, 2006a). The group mean O_3- induced FEV_1 decrements observed in these studies were 3.6% by Adams (1998), 2.8% (triangular exposure) and 2.9% (S-W exposure) by Adams (2006a), 3.5% by Schelegle *et al.* (2009), and 1.8% by Kim *et al.* (2011). Based on data from these four studies, at 60 ppb, the weighted-average group mean O_3-induced FEV_1 decrement (*i.e.*, adjusted for FA responses) is 2.7% (n=150) (Kim *et al*., 2011; Schelegle *et al.*, 2009; Adams, 2006a, 1998). Although ***not consistently statistically significant***, these group mean changes in FEV_1 at 60 ppb are consistent between studies, *i.e.*, none observed an average improvement in lung function following a 6.6-h exposure to 60 ppb O_3. [emphasis added]

The group mean change in FEV_1 at 60 ppb was only statistically significant in the study by Kim *et al*. (2011); thus, the findings from the other studies examining this exposure level are consistent with there being no effect. Despite this, much has been made of the results of the Adams (2006) study. It has been argued recently by some Clean Air Scientific Advisory Committee (CASAC) members that Adams (2006) only found statistically non-significant results because the statistical test, specifically the Scheffe *post hoc* test, was not sufficiently powerful to detect the effect. The Scheffe test used by Adams (2006) is a commonly used statistical test to compare mean values that minimizes false positives, but may be more likely to produce false negatives. Some CASAC members suggested that this test is overly conservative and had other statistical tests been used, results would have been statistically significant (Allen *et al*., 2011).

Brown (2007) and Brown *et al*. (2008) claim to address this by analyzing only the 6.6-hour response at 60 ppb ozone *vs*. FA in the Adams (2006) study. This approach excluded all pulmonary function data at other interim hourly time points and exposure levels within the 6.6-hour exposure pattern. This approach is also at variance with those of other research groups that have performed prolonged ozone exposures and published their results in the scientific literature prior to the Brown reanalysis, including those by researchers at the University of Rochester (Torres *et al*., 1997), the University of Toronto (Liu *et al*., 1999), the University of California, Los Angeles (Gong *et al*., 1998), and EPA (Gong *et al*., 2004). In a presentation to CASAC, Professor William Adams expressed concerns with EPA's

reanalysis of selected data from his study and its conclusions, which are very different from those in the published paper (Adams, 2007). While the approach used by Brown (2007) and Brown *et al*. (2008) produced statistically significant results, this can be attributed to the majority of the data being selectively omitted from the analysis.

Dr. Mark Nicolich conducted an analysis of the full data set from Adams (2006) using a mixed model analysis of variance and Dunnett's *post hoc* test instead of the Scheffe test (Nicolich, 2007). This reanalysis, using a technique that is less likely to produce false negatives, was consistent with the original finding by Adams (2006) that there was no statistically significant decrement in group mean FEV_1 after exposure to 60 ppb ozone *versus* FA after 6.6 hours of exercise. In addition, Lefohn *et al*. (2010) re-analyzed five controlled ozone exposure studies, including those by Adams (2006) and Schelegle *et al*. (2009), using two-factor analysis of variance (ANOVA) and evaluating statistical significance using the Tukey's studentized range approach to account for multiple comparisons for least square means. Although they did not subtract the FA FEV_1 from the ozone-treatment responses, their methodology can still be considered conservative in that it minimizes Type II errors. They did not find any statistically significant changes in FEV_1 at any measurement time associated with the 40 ppb and 60 ppb profiles. In addition, they found that in four out of five studies, exposures to FA substantially improved FEV_1 response over the exposure period, which means that analyses using FA controls likely overestimated FEV_1 changes.

EPA has given no scientifically acceptable justification for relying on the Brown statistical analyses over the original analyses conducted by the authors or those of Nicolich (2007) or Lefohn *et al*. (2010). While each statistical method has strengths and limitations, several scientifically accepted statistical methods indicate that there is no statistically significant association between exposure to 60 ppb ozone and lung function decrements. We recommend that in the ISA, EPA recognize and give equal or greater weight to analyses using these methods, as well as to other methods that incorporate all of the exposure concentrations and time points, such as meta-regression.

3.3 EPA Does Not Fully Consider Adversity Criteria

Section 109 of the Clean Air Act directs the EPA Administrator to set and revise a primary NAAQS "to protect against adverse health effects" of criteria pollutants. EPA does not fully consider the

criteria for determining the adversity of health effects associated with controlled ozone exposure in the First and Second Draft ISA, however.

There is no indication that the reported FEV_1 decrements at 60 ppb in the controlled human exposure studies are adverse, regardless of whether they are statistically significant. Regarding what constitutes an adverse effect on pulmonary function, we recommend that EPA add the text from the American Thoracic Society (ATS) publication that states:

> The committee recommends that a small, transient loss of lung function, by itself should not automatically be designated as adverse. In drawing the distinction between adverse and nonadverse reversible effects, this committee recommended that reversible loss of lung function in combination with the presence of symptoms should be considered adverse. (ATS, 2000)

EPA should also note that the European Respiratory Society (ERS) suggests that only short-term changes in FEV_1 exceeding 12% "may be clinically important" (Pellegrino *et al*., 2005).

EPA has concluded previously that a moderate decrement in lung function (defined as a decrease in FEV_1 between 10% and 20%) *and/or* respiratory symptoms are considered to be adverse (US EPA, 2007). We recommend that EPA clearly indicate that their criterion for adversity differs from that of the ATS, as the ATS recommendation does not consider small changes in lung function alone to be adverse. Also, we note that ATS states that EPA's criterion of 10% has not been validated for acceptability or against other measures (ATS, 2000). We recommend that EPA clarify in the ISA that the 10% criterion for adversity is EPA's own internal criterion and is not based on ATS guidelines because ATS has not provided specific numeric criteria for changes in FEV_1 or other measures of lung function.

Regardless, the FEV_1 decrements reported at 60 ppb ranged from 1.7-3.5% and were not accompanied by an increase in respiratory symptoms. EPA notes in the Second Draft ISA that changes in FEV_1 measurement should exceed 5% to overcome the intra-day variability of FEV_1 in normal subjects (Pellegrino *et al*., 2005). In addition, the decrements were transient, reversible, and of low severity in that they did not interfere with normal activity and would not result in permanent respiratory injury or progressive respiratory dysfunction. Although some individuals had larger decrements, as discussed below, these cannot be attributed to ozone. Overall, the lung function effects at 60 ppb ozone in controlled exposure studies are within the range of intra-individual variability in normal subjects and are not considered adverse with respect to broadly recognized clinical guidelines (*e.g.*, ATS and ERS). The

lowest ozone concentration associated with both an FEV_1 decrement >10% and increased respiratory symptoms, which is considered an adverse effect based on clinical guidelines, is 88 ppb, as reported in the study by Schelegle *et al*. (2009).

Another example of how EPA does not fully consider adversity criteria is seen in the Preamble of the Second Draft ISA, in which EPA discusses the official statement by the ATS (2000) that suggests that if the relationship between a risk factor and a disease is causal, the shift in the risk factor distribution of the exposed population, and hence the risk profile of the exposed population, should be considered adverse, even in the absence of the immediate occurrence of frank illness. In order to determine whether this criterion is met, one must rely on studies from which it is appropriate to extrapolate to the general population. While the controlled exposure studies provide a conservative estimate of ozone risks (because the enhanced exercise regimens in these studies compensate for compromised respiratory systems in sensitive populations), it is not appropriate to use these studies to determine at what exposure there will be a population shift in FEV_1 decrements.

The controlled exposure study subjects were not selected to be representative of the general US population. They were young, healthy adults who exercised during the study. Ozone responsiveness is highest in young adults; it has been demonstrated to be halved in children (McDonnell *et al*., 1985) and, after young adulthood, to decrease with increasing age (McDonnell *et al*., 1995, 2007). Lung function measurements in these individuals are not applicable to asthmatics or other sensitive people who are not able to achieve the total cumulative absorbed ozone doses in the controlled exposure studies because of their inability to exercise strenuously for long durations. For example, while the nominal ambient ozone concentration in the Adams (2006) study at issue is 60 ppb ozone, the subjects were exercising vigorously (V_e = ~20 L/min/m^2 body surface area) over 6.6 hours, so the absorbed dose of ozone was substantially higher than it would have been were the subjects at rest, and also likely higher than it would be in average workers who conduct manual labor (see Linn *et al*., 1993). Indeed, the absorbed dose can probably be safely said to be much higher than that of sensitive populations exerting even moderate physical effort. Because this is the case, Adams (2006) and other studies conducted with high ventilation rates have a bias toward overestimating the effect on lung function response, providing a margin of safety for sensitive individuals in the population.

In addition, the number of subjects in the controlled exposure studies is far too small to estimate the proportions of individuals that would be affected in the general population. In this vein, it is not

appropriate to use frequency distributions (as EPA did in Figure 6-2 of the Second Draft ISA) to assess population shifts in FEV_1 decrements. This is because the proportion of study subjects with FEV_1 decrements ≥10% is too uncertain to determine the proportion that would occur in the general population. Furthermore, these proportions are based on one measurement per individual, so they do not take into account intra-individual variability. Given that intra-individual variability is on the order of 5%, it is highly likely that some individual responses are misclassified.

In the ISA, EPA suggests that small changes in FEV_1 that are claimed to be statistically significant but are not clinically relevant are indicative of an adverse effect of ozone, presumably because these small changes indicate a population shift in the risk profile of the exposed population. The controlled exposure studies are useful in that they provide a conservative estimate of ozone risks, but they are not designed to determine at what exposure level a population shift in effects occurs. If EPA elects to continue using the results of chamber studies with a small number of subjects to make broad conclusions about large populations, we recommend that the ISA include a clarification that this approach is at variance with ATS criteria. The ATS (2000) guidelines clearly draw a distinction between small changes in lung function alone in individuals, as derived from clinical studies (which ATS does not consider to be adverse), and small but significant reductions in the lung function of a population, as derived from observational epidemiology studies (which ATS considers to be clinically significant).

3.4 Controlled Human Exposure Studies Were Not Designed to Assess Individual Data

All of the controlled human exposure studies were designed to assess differences on the group mean level, not changes in a specific individual. Because of this, one cannot determine whether effects in certain individuals, even if quite large, are in fact representative of any generalized effect of ozone exposure. If one were interested in determining whether changes in lung function in a given individual were due to ozone, one would have to design a study that has several repeat measurements for each individual, and perform a scientifically acceptable statistical test on the data for each individual. Otherwise one cannot know whether individual changes are due to ozone, chance, or some other factor (Pagano and Gauvreau, 2000).

The misleading nature of over interpreting individual findings is illustrated by a review of the data from the Kim *et al*. (2011) and Adams (2006) studies. Kim *et al*. (2011) reported three people with

FEV_1 decrements >10% at 60 ppb. Two of these people also underwent 80 ppb exposures, but had a lesser FEV_1 decrement at 80 ppb than at 60 ppb. Similarly, in the study by Adams (2006), one of two individuals with the square-wave exposures and one of two with triangular exposures also had lower FEV_1 decrements after exposure to 80 ppb than to 60 ppb. Because each of these individuals only had one measurement at 60 ppb and 80 ppb, respectively, at the end of the 6.6-hour exposure protocol (when the statistical test was conducted), one cannot determine, ***based on the data***, whether or not 80 ppb is actually healthier, the same, or worse than 60 ppb in each of these individuals. Because it is biologically implausible that 80 ppb is generally healthier than 60 ppb, and in fact EPA and CASAC seem to conclude that 80 ppb less healthy than 60 ppb, the anomalous data from these individuals suggests that the individual data from these studies are not a reliable measure of general population response.

In an evaluation of the Adams (2006) and four other controlled exposure studies, Lefohn *et al.* (2010) stated, "Additional common findings, based on our reanalysis, among healthy exercising young adults included (i) high intraindividual variability in subject response within exposure profiles; (ii) inconsistent individual FEV_1 response patterns across exposure profiles; [and] (iii) FA exposure FEV_1 changes up to ± 5% in some subjects." This indicates that any FEV_1 decrement (or increment) should have an error bar of at least 5% and that a large proportion of most measured FEV_1 responses must be attributable to factors other than ozone in these studies.

On pages 6-12 to 6-14 of the Second Draft ISA, EPA presents a discussion on inter-subject variability in responses of healthy subjects. To facilitate this discussion, EPA presents "frequency distributions" of FEV_1 decrements from the study by Schelegle *et al.* (2009) in Figure 6-2. In the discussion and distribution charts, EPA fails to recognize the variation attributable to experimental factors; rather, EPA assigns all variability to individual response differences. We strongly recommend that EPA include error bars on the distribution charts, and include in the discussion that at low ozone exposure concentrations (*e.g.*, 60 ppb), the magnitude of the experimental error and daily variation in FEV_1 is significant compared to the small changes possibly due to test exposure.

As noted above, EPA considers an FEV_1 decrement > 10% to be adverse. The controlled human exposure studies are informative regarding the exposure level, on average, at which group mean FEV_1 decrements differ from those in FA. They are not informative regarding whether responses in any particular individual are fully caused by ozone, however, even if measured FEV_1 decrements are > 10%.

Studies that were not designed to determine whether effects in individual subjects are attributable to ozone should not be reanalyzed or interpreted as such in the ISA.

4 Conclusions

Gradient previously identified several issues with EPA's approach for evaluating controlled human exposure, short-term epidemiology, and long-term epidemiology studies of ozone and the causal determinations that were made in the First Draft ISA for ozone. These issues are still evident in the Second Draft ISA.

In EPA's evaluation of epidemiology data, there are no clear statements indicating how the evidence is weighed, how evaluations of the various Bradford Hill aspects are to be carried out, how judgments regarding whether aspects are to be considered as met are made, and how an overall judgment is to be synthesized out of consideration of all of the aspects jointly. EPA also does not follow through with weighing alternative views on controversial issues. Thus, the Second Draft ISA does not provide an objective appraisal of the epidemiology evidence.

EPA's evaluation of the controlled human exposure data is not scientifically appropriate. EPA states that there is no evidence of a threshold for lung function effects of ozone, and that effects at exposures of 60 ppb are statistically significant and adverse. EPA also over interpreted data from the controlled exposure studies to determine whether changes in lung function in a given individual were attributable to ozone. The controlled exposure studies indicate a threshold below 70 ppb at which there are no statistically significant adverse effects associated with ozone, and this is consistent with biological data that support a threshold mechanism of action. Effects at 60 ppb are not adverse, nor do they occur statistically more often than do those associated with FA exposures. In addition, any large decrements in lung function in particular individuals at this exposure level cannot be attributed to ozone.

Overall, the Second Draft ISA does not follow an objective and rigorous weight-of evidence approach for evaluating the available data and does not evaluate the controlled human exposure data in a scientifically appropriate manner. Because of this, the ISA cannot be relied upon to support a causal relationship between ozone exposure and health outcomes at exposure levels below the current NAAQS. EPA should follow the NRC recommendations for best practices in risk assessment to ensure its goal of providing credible causal conclusions about ozone-related health effects in the final ISA.

References

Adams, WC. 2002. "Comparison of chamber and face-mask 6.6-hour exposures to ozone on pulmonary function and symptoms responses." *Inhal. Toxicol.* 14(7):745-764.

Adams, WC. 2003. "Comparison of chamber and face mask 6.6-hour exposure to 0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses." *Inhal. Toxicol.* 15:265-281.

Adams, WC. 2006. "Comparison of chamber 6.6-h exposures to 0.04-0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses." *Inhal. Toxicol.* 18(2):127-136.

Adams, WC. 2007. "Public Comment to CASAC Ozone Review Panel Teleconference." University of California, Davis, CA, 2p., March 5.

Allen, G; Balmes, J; Gauderman, J; Henderson, R; Hopke, PK; Sheppar, L; Speizer, F; Suh, H. 2011. "Preliminary Individual Comments on the Ozone Reconsideration." 24p., February 14.

American Thoracic Society (ATS). 2000. "What constitutes an adverse health effect of air pollution?" *Am. J. Respir. Crit. Care Med.* 161:665-673.

Brown, JS. [US EPA]. 2007. Memo to Ozone NAAQS Review Docket (OAR-2005-0172) re: The effects of ozone on lung function at 0.06 ppm in healthy adults. 8p., June 14.

Brown, JS; Bateson, TF; McDonnell, WF. 2008. "Effects of exposure to 0.06 ppm ozone on FEV_1 in humans: A secondary analysis of existing data." *Environ. Health Perspect.* 116(8):1023-1026.

Folinsbee, LJ; McDonnell, WF; Horstman, DH. 1988. "Pulmonary function and symptom responses after 6.6-hour exposure to 0.12 ppm ozone with moderate exercise." *J. Air Waste Manag. Assoc.* 38:28-35.

Frey, HC; Samet, JM. 2011. Draft letter to The Honorable Lisa P. Jackson re: CASAC review of EPA's Integrated Science Assessment for Lead (First External Review Draft – May 2011). October 24, 2011. Available at: http://yosemite.epa.gov/sab/sabproduct.nsf/ea5d9a9b55cc319285256cbd005a472e/48ee1b8614c830fe85257937005be49c/$FILE/CASAC_Lead_Review_Panel_1st_ERD_ISA_Letter-10-25-11.pdf

Girardot, SP; Ryan, PB; Smith, SM; Davis, WT; Hamilton, CB; Obenour, RA; Renfro, JR; Tromatore, KA; Reed, GD. 2006. "Ozone and PM2.5 exposure and acute pulmonary health effects: a study of hikers in the Great Smoky Mountains National Park." *Environ Health Perspect.* 114(7):1044-52.

Gong, H; Linn, W; Terrel, S; Anderson, K; Clark, K; Sioutas, C; Casio, W; Alexis, N; Devlin, R. 2004. "Exposure of elderly volunteers with and without chronic obstructive pulmonary disease to concentrated ambient fine particulate pollution." *Inhalation Toxicol.* 16:731-744.

Gong, H; Wong, R; Sarma, RJ; Linn, WS; Sullivan, ED; Shamoo, DA; Anderson, KR; Prasad, SB. 1998. "Cardiovascular effects of ozone exposure in human volunteers." *Am. J. Respir. Crit. Care Med.* 158(2):538-546.

Goodman, JE. 2011. "Comments on First Draft Ozone Integrated Science Assessment." Submitted to Docket No. EPA-HQ-ORD-2011-0050, 38p., May 2.

Horstman, DH; Folinsbee, LJ; Ives, PJ; Abdul-Salaam, S; McDonnell, WF. 1990. "Ozone concentration and pulmonary response relationships for 6.6-hour exposures with five hours of moderate exercise to 0.08, 0.10, and 0.12 ppm." *Am. J. Respir. Crit. Care Med.* 142:1158-1163.

Kim, CS; Alexis, NE; Rappold, AG; Kehrl, H; Hazucha, MJ; Lay, JC; Schmitt, MT; Case, M; Devlin, RB; Peden, DB; Diaz-Sanchez, D. 2011. "Lung function and inflammatory responses in healthy young adults exposed to 0.06 ppm ozone for 6.6 hours." *Am. J. Respir. Crit. Care Med.* 183:1215-1221..

Korrick, SA; Neas, LM; Dockery, DW; Gold, DR; Allen, GA; Hill, LB; Kimball, KD; Rosner, BA; Speizer, FE. 1998. "Effects of ozone and other pollutants on the pulmonary function of adult hikers." *Environ. Health Perspect.* 106(2):93-99.

Lefohn, AS; Hazucha, MJ; Shadwick, D; Adams, WC. 2010. "An alternative form and level of the human health ozone standard." *Inhal. Toxicol.* 22(12):999-1011.

Linn, WS; Spier, CE; Hackney, JD. 1993. "Activity patterns in ozone-exposed construction workers." *J. Occup. Med. Tox.* 2:1-14.

Liu, L; Leech, JA; Urch, RB; Poon, R; Zimmerman, B; Kubay, JM; Silverman, FS. 1999. "A comparison of biomarkers of ozone exposure in human plasma, nasal lavage and sputum." *Inhal. Toxicol.* 11(8):657-674.

McDonnell, WF; Chapman, RS; Leigh, MW; Strope, GL; Collier, AM. 1985. "Respiratory responses of vigorously exercising children to 0.12 ppm ozone exposure." *Am. Rev. Respir. Dis.* 132:875-879.

McDonnell, WF; Kehrl, HR; Abdul-Salaam, S; Ives, PJ; Folinsbee, LJ; Devlin, RB; O'Neil, JJ; Horstman, DH. 1991. "Respiratory response of humans exposed to low levels of ozone for 6.6 hours." *Arch. Environ. Occup. Health.* 46:145-150.

McDonnell, WF; Stewart, PW; Andreoni, S; Smith, MV. 1995. "Proportion of moderately exercising individuals responding to low-level, multi-hour ozone exposure." *Am. J. Respir. Crit. Care Med.* 152:589 596.

McDonnell, WF; Stewart, PW; Smith, MV. 2007. "The temporal dynamics of ozone-induced FEV_1 changes in humans: An exposure-response model." *Inhal. Toxicol.* 19(6-7):483-494.

Mortimer, KM; Neas, LM; Dockery, DW; Redline, S; Tager, IB. 2002. "The effect of air pollution on inner-city children with asthma." *Eur. Respir. J.* 19:699-705.

National Research Council (NRC). 2011. "Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde (Prepublication Copy)." National Academies Press, Washington, DC, 155p. Accessed on April 8, 2011 at http://www.nap.edu/catalog/13142.html, April.

Nicolich, M. 2007. Letter Report to US EPA, Air and Radiation Docket re: Comments on EPA's Proposed "National Ambient Air Quality Standards for Ozone," 72 *Fed Reg*. 37,818 (July 11, 2007). Docket No. EPA-HQ-OAR-2005-0172, October 9.

O'Connor, GT; Neas, L; Vaughn, B; Kattan, M; Mitchell, H; Crain, EF; Evans, R III; Gruchalla, R; Morgan, W; Stout, J; Adams, GK; Lippman, M. 2008. "Acute respiratory health effects of air pollution on children with asthma in US inner cities." *J. Allergy Clin. Immunol.* 121(5):1133-1139.

Pagano, M; Gauvreau, K. 2000. *Principles of Biostatistics - Second Edition*. Duxbury, Pacific Grove, CA, 525p.

Pellegrino, R; Viegi, G; Brusasco, V; Crapo, RO; Burgos, F; Casaburi, R; Coates, A; van der Grinten, CP; Gustafsson, P; Hankinson, J; Jensen, R; Johnson, DC; MacIntyre, N; McKay, R; Miller, MR; Navajas, D; Pedersen, OF; Wanger, J. 2005. "Interpretative strategies for lung function tests." *Eur. Respir. J.* 26(5):948-968.

Prueitt, RL; Goodman, JE; Bailey, LA; Rhomberg, LR. 2011. "Hypothesis-based weight-of-evidence evaluation of the neurodevelopmental effects of chlorpyrifos." *Crit. Rev. Toxicol.* In press.

Rhomberg, LR; Bailey, LA; Goodman, JE. 2010. "Hypothesis-based weight of evidence: A tool for evaluating and communicating uncertainties and inconsistencies in the large body of evidence in proposing a carcinogenic mode of action - Naphthalene as an example." *Crit. Rev. Toxicol.* 40:671-696.

Rhomberg, LR; Bailey, LA; Goodman, JE; Hamade, A; Mayfield, D. 2011. "Is exposure to formaldehyde in air causally associated with leukemia? - A hypothesis-based weight-of-evidence analysis." *Crit. Rev. Toxicol.* 41(7):555-621.

Schelegle, ES; Morales, CA; Walby, WF; Marion, S; Allen, RP. 2009. "6.6-Hour inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans." *Am. J. Respir. Crit. Care Med.* 180(3):265-272.

Schildcrout, JS; Sheppard, L; Lumley, T; Slaughter, JC; Koenig, JQ; Shapiro, GG. 2006. "Ambient air pollution and asthma exacerbations in children: An eight-city analysis." *Am. J. Epidemiol.* 164(5):505-517.

Torres, A; Utell, MJ; Morow, PE; Voter, KZ; Whitlin, JC; Cox, C; Looney, RJ; Spers, DM; Tsai, Y; Frampton, MW. 1997. "Airway inflammation in smokers and nonsmokers with varying responsiveness to ozone." *Am. J. Respir. Crit. Care Med.* 156:728-736.

US EPA. 2007. "Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information (OAQPS Staff Paper)." Office of Air Quality Planning and Standards, EPA-452/R-07-007, 609p., July.

US EPA. 2011a. "Integrated Science Assessment for Ozone and Related Photochemical Oxidants (Second External Review Draft)." National Center for Environmental Assessment (NCEA), EPA/600/R-10/076B, 1222p., September.

US EPA. 2011b. "Integrated Science Assessment for Ozone and Related Photochemical Oxidants (First External Review Draft)." National Center for Environmental Assessment (NCEA), EPA/600/R 10/076A, 996p., March.

US EPA. 2011c. "Weight of Evidence: Evaluating Results of EDSP Tier 1 Screening to Identify the Need for Tier 2 Testing." Office of Chemical Safety and Pollution Prevention, Endocrine Disruptor Screening Program, 47p., September.

Appendix A

Data for Evaluation of an Ozone Concentration Threshold for Pulmonary Function Effects After Acute Exposure in Healthy Subjects

Table 1
Data for Evaluation of an Ozone Concentration Threshold for Pulmonary Function Effects After Acute Exposure in Healthy Subjects

Study	Exposure	No. of Subjects (N)	O_3 Concentration (ppb)	ΔFEV_1 (%)[1]	Source	Standard Error (SE)	Source	Standard Deviation (SD)[2]	Source	Group Decrease in FEV_1 (%)[3]	Standard Deviation (SD)[4]	Standard Error (SE)[5]
		31	Air	0.8	From the Study	0.9	From the Study	5.011	Calculated from SE			
Schelegle et al. (2009)	CH-TRI	31	63	-2.72	From the Study	1.48	From the Study	8.240	Calculated from SE	3.52	8.31	1.49
	CH-TRI	31	72	-5.34	From the Study	1.42	From the Study	7.906	Calculated from SE	6.14	8.03	1.44
	CH-TRI	31	81	-7.02	From the Study	1.6	From the Study	8.908	Calculated from SE	7.82	8.86	1.59
	CH-TRI	31	88	-11.42	From the Study	2.2	From the Study	12.249	Calculated from SE	12.22	11.81	2.12
Folinsbee et al. (1988)		10	Air	1.9	From the Study			6	From the Study			
	CH-SQR	10	120	-13	From the Study			15.4	From the Study	14.9	14.81	4.68
Horstman et al. (1990)		22	Air	0.6	From the Study			NA				
	CH-SQR	22	80	-7	From the Study			NA		7.6		
	CH-SQR	22	100	-7	From the Study			NA		7.6		
	CH-SQR	21	120	-12.3	From the Study			NA		12.9		
McDonnell et al. (1991)		38	Air	-0.7	Absolute change in the Study; Converted to %			0.66	Absolute change in the Study; Converted to %			
	CH-SQR	38	80	-8.4	Absolute change in the Study; Converted to %			1.76	Absolute change in the Study; Converted to %	7.75	1.69	0.27
		10	Air	2.4	Absolute change in the Study; Converted to %			15.42	Absolute change in the Study; Converted to %			
	CH-SQR	10	100	-11.4	Absolute change in the Study; Converted to %			4.60	Absolute change in the Study; Converted to %	13.80	14.75	4.66
Adams (2003)	CH	30	Air	2.65	From the Study			3.43	From the Study			
	CH-SQR	30	80	-3.51	From the Study			7.43	From the Study	6.16	7.22	1.32
	CH-TRI	30	80	-3.12	From the Study			6.08	From the Study	5.77	6.05	1.10
	FM	30	Air	2.5	From the Study			3.61	From the Study			
	FM-SQR	30	80	-3.64	From the Study			7.8	From the Study	6.14	7.58	1.38
	FM-TRI	30	80	-2.95	From the Study			5.58	From the Study	5.45	5.70	1.04
Adams (2002)	CH	30	Air	2.39	From the Study			4.01	From the Study (Assumed to be SD)			
	CH-SQR	30	120	-13.25	From the Study			11.19	From the Study (Assumed to be SD)	15.64	10.73	1.96
	FM-SQR	30	40	1.15	From the Study			4.2	From the Study (Assumed to be SD)	1.24	4.89	0.89
	FM-SQR	30	80	-3.96	From the Study			7.5	From the Study (Assumed to be SD)	6.35	7.40	1.35
	FM-SQR	30	120	-13.02	From the Study			9.21	From the Study (Assumed to be SD)	15.41	8.91	1.63
Adams (2006)	SQR	30	Air	1.35	From the Study			2.98	From the Study (Assumed to be SD)			
	CH-SQR	30	80	-4.72	From the Study			8.65	From the Study (Assumed to be SD)	6.07	8.29	1.51
	CH-SQR	30	60	-1.51	From the Study			4.24	From the Study (Assumed to be SD)	2.86	4.42	0.81
	CH-TRI	30	80	-5.65	From the Study			8.08	From the Study (Assumed to be SD)	7.00	7.76	1.42
	CH-TRI	30	60	-1.43	From the Study			5.95	From the Study (Assumed to be SD)	2.78	5.83	1.06
	CH-TRI	30	40	1.17	From the Study			2.97	From the Study (Assumed to be SD)	0.18	3.54	0.65
Kim et al. (2011)		59	Air	-0.002	From the Study	0.46	From the Study	3.53	Calculated from SE			
	CH-SQR	59	60	-1.71	From the Study	0.50	From the Study	3.84	Calculated from SE	1.71	4.40	0.57

Notes:
NA = Not Available
[1] Percent change in Forced Expiratory Volume in One Second
[2] When Standard Error (SE) were used to calculated Standard Deviation (SD), $SD = SE \times \sqrt{N}$
[3] Group Decrease in $FEV_1 = \Delta FEV_{1,Air} - \Delta FEV_{1,O3}$
[4] $SD = \sqrt{[SD(\Delta FEV_{1,Air})]^2 + [SD(\Delta FEV_{1,O3})]^2 - 2\times\rho\times SD(\Delta FEV_{1,Air})\times SD(\Delta FEV_{1,O3})}$; where ρ (=0.291) is the average correlation coefficient for $\Delta FEV_{1,O3}$ versus $\Delta FEV_{1,Air}$ for individuals [calculated from individual subject data in Kim *et al*. (2011) and Adams (2006) studies, see Figure A-1).
[5] $SE = SD/\sqrt{N}$

Appendix B

Gradient's Comments on
First Draft Ozone Integrated Science Assessment

Comments on First Draft Ozone Integrated Science Assessment

Julie E. Goodman, Ph.D., DABT

Sponsored by
American Petroleum Institute
1220 L Street NW
Washington DC 20005-4070

May 2, 2011

Gradient

Table of Contents

Page

Executive Summary ES-1

1 Introduction 1

2 Controlled Exposure Studies 2
- 2.1 Controlled Exposure Study Protocols 2
- 2.2 A Cross-Study Comparison Indicates a Threshold Below 70 ppb Ozone at Which There are No Significant Adverse Effects 4
- 2.3 Antioxidant Defenses Against Ozone Indicate a Threshold Mode of Action 5
- 2.4 Mean FEV_1 Decrements at 60 ppb are Not Statistically Significant 6
- 2.5 Controlled Human Exposure Studies were Not Designed to Assess Individual Data 8
- 2.6 Mean FEV_1 Decrements at 60 ppb are Not Adverse 9
- 2.7 One Cannot Determine a Population Shift in FEV_1 Changes from Controlled Exposure Studies Conducted with Subjects with High Ventilation Rates 10

3 Epidemiology Studies 13
- 3.1 Short-Term Respiratory Morbidity 13
  - 3.1.1 EPA Frequently Misinterprets or Discounts Evidence from Studies Reporting Null Results 13
  - 3.1.2 EPA Does Not Adequately Consider Factors that Can Bias Study Results 15
  - 3.1.3 EPA Does Not Appropriately Weigh the Evidence 16
  - 3.1.4 Conclusions for Short-term Respiratory Morbidity 16
- 3.2 Short-Term Cardiac Morbidity 17
- 3.3 Short-Term Mortality 18
  - 3.3.1 EPA Does Not Adequately Acknowledge that the Relationship Between Short-term Ozone Exposure and Mortality is Significantly Confounded by Other Pollutants 19
  - 3.3.2 EPA Does Not Adequately Consider the Sensitivity of Mortality Risk Estimates to Model Specification 19
  - 3.3.3 EPA Does Not Adequately Consider the Unexplained Regional Heterogeneity in Ozone-Mortality Risk Estimates 20
  - 3.3.4 EPA Inappropriately Interprets Mixed Evidence for an Ozone Concentration-Response Threshold as No Evidence to Support a Threshold 20
  - 3.3.5 EPA Does Not Provide Information Regarding How It Weighs the Evidence for Causation 21
  - 3.3.6 Conclusions for Short-Term Mortality 22
- 3.4 Long-Term Respiratory Morbidity 22
- 3.5 Long-Term Mortality 23

References 26

Appendix A

Executive Summary

In March, 2011, the United States Environmental Protection Agency (EPA) released the first external review draft of the *Integrated Science Assessment for Ozone and Related Photochemical Oxidants*, or "ISA." In the ISA, EPA evaluated controlled human exposure, short-term epidemiology, and long-term epidemiology studies of ozone and made causal determinations regarding the relationship between ozone exposure and the health effects examined in these studies. There are many issues associated with EPA's evaluation. A scientifically rigorous analysis does not support a causal relationship between ozone exposure and respiratory morbidity, cardiovascular (CV) morbidity, or mortality at exposure levels below the current National Ambient Air Quality Standards (NAAQS).

Controlled exposure studies that assessed the association between ozone and lung function at exposures below 80 ppb indicate a threshold below 70 ppb at which there are no statistically significant adverse effects associated with ozone, and this is consistent with biological data that support a threshold mechanism of action. Effects at 60 ppb are not adverse with respect to broadly recognized clinical guidelines, nor do they occur statistically more often than do those associated with filtered air (FA) exposures. Furthermore, any large decrements in lung function in particular individuals at this exposure level cannot be attributed to ozone. In this vein, one cannot determine whether there would be a population shift in one-second forced expiratory volume (FEV_1) decrements at 60 ppb ozone exposure based on the small number of individuals in the controlled exposure studies.

Recent studies that examined associations between short-term ambient ozone exposure and respiratory morbidity have reported mixed results across all health endpoints. EPA's approach to assessing this evidence in the ISA is inappropriate and does not consider many of the limitations and uncertainties associated with the underlying studies. EPA focuses on the positive associations reported in these studies, whether they are statistically significant or not, and often discounts null results. EPA does not adequately consider the factors that can bias study results, such as exposure or outcome measurement error and the choice of lag period. EPA also does not appear to appropriately weigh the evidence for causality, and provides summarizing statements that emphasize only positive associations. A scientifically valid assessment of the evidence for short-term respiratory morbidity does not support a causal association.

EPA bases their "suggestive" causality conclusion for short-term ambient ozone exposure and CV morbidity on the evidence from high-dose animal toxicology studies, as well as "consistent" evidence of an association between ozone and CV mortality, but notes the weak coherence and biological plausibility for ozone-induced CV morbidity. Inconclusive evidence from epidemiology studies and very high-dose animal studies is not sufficient to conclude there is a "suggestive" causal relationship between short-term ozone exposure and CV morbidity, and it appears that the conclusions made by EPA regarding causation are based on a subjective view of the overall data.

There are multiple uncertainties in the assessment of the ozone-mortality relationship, such as confounding by various forms of particulate matter (PM), sensitivity of ozone time-series models to model specification, and unexplained regional heterogeneity. The recent studies reviewed in the ISA provide evidence for these sources of uncertainty, but EPA does not adequately address the uncertainties in the ISA. EPA also does not appropriately weigh the evidence for cause-specific mortality, which indicates a large number of null results. In addition, the evidence for an association between ozone and acute CV morbidity provides a marked lack of coherence for an association between short-term ozone exposure and CV mortality. We recommend that EPA change its causality conclusion to reflect the fact that the current evidence is insufficient to establish a causal relationship between short-term ozone exposure and mortality.

The determination of the relationship between long-term ozone exposure and respiratory morbidity in the ISA as "likely to be causal" is questionable. The recent evidence cited for this causal determination mainly comes from studies of asthma-related outcomes in children with specific gene variants, but these studies do not demonstrate any consistent associations between ozone exposure and these outcomes. The long-term studies of respiratory morbidity suffer from many of the same limitations as the short-term studies, such as exposure measurement error from the use of central monitors and confounding by other pollutants, and these limitations are not adequately addressed in the ISA. Overall, recent evidence does not provide support for a "likely to be causal relationship," and it is unclear how this evidence is more compelling than that of other health outcomes for which the evidence was determined by EPA to only be "suggestive" of a causal relationship.

In the assessment of the evidence regarding whether there is an association between long-term ozone exposure and mortality, EPA cites multiple studies that reported null results, as well as two studies conducted since the last review (one of which reported null results) as new evidence. EPA also states that results of short- and long-term respiratory morbidity studies provide biological plausibility for mortality

due to respiratory disease, but then concludes that the collective evidence is suggestive of a causal relationship between long-term ozone exposure and *all-cause* mortality, not just respiratory mortality. EPA acknowledges that the available data regarding respiratory and cardio-pulmonary mortality show no association, with the exception of one study by Jerrett *et al*. (2009) reporting an association with respiratory-specific mortality. The many limitations of the Jerrett *et al*. (2009) study weaken the evidence for respiratory-specific mortality, however. A casual relationship is also not coherent with the overall inconsistent evidence for respiratory morbidity and the almost exclusively null evidence for short-term effects on respiratory mortality. Together, the evidence for an association between long-term exposure to ozone and mortality is overwhelmingly null, and one weakly positive study with many limitations is not sufficient to suggest a causal relationship.

1 Introduction

In March, 2011, the United States Environmental Protection Agency (US EPA, 2011) released the first external review draft of the *Integrated Science Assessment for Ozone and Related Photochemical Oxidants* (referred to as "the ISA" throughout these comments). In the ISA, EPA evaluated controlled human exposure, short-term epidemiology, and long-term epidemiology studies of ozone and made causal determinations regarding the relationship between ozone exposure and the health effects examined in these studies. There are several issues associated with EPA's approach for evaluating human results in the ISA. These include:

- Insufficient documentation of methods and criteria for identifying the epidemiologic evidence to be reviewed;
- Insufficient description of methods for evaluating the strengths and limitations of the selected studies, leading to inconsistent evaluation of individual studies;
- The lack of a clear framework for evaluating the weight of the evidence in establishing causation, leading to inconsistent, *ad hoc* approaches to hazard identification; and
- The lack of a definition of the concept of consistency of findings within and among studies, leading to conclusions about causation that appear to be based on a subjective view of the overall data.[1]

Below, we evaluate EPA's assessment of the evidence regarding respiratory morbidity, cardiovascular (CV) morbidity, and mortality as presented in the ISA. For short-term ozone exposure, EPA concludes that a causal relationship exists for respiratory morbidity, there is likely to be a causal relationship with mortality, and that the evidence is suggestive of a causal relationship for CV morbidity (US EPA, 2011). For long-term ozone exposure, EPA concludes that there is likely to be a causal relationship with respiratory morbidity, and the evidence is suggestive of a causal relationship for mortality. Because of the issues noted above, the evidence for these outcomes is inappropriately evaluated in the ISA. A scientifically valid analysis does not support a causal relationship between ozone exposure and these health outcomes at exposure levels below the current National Ambient Air Quality Standards (NAAQS).

[1] In 2011, a committee assembled by the National Research Council (NRC) of the National Academy of Sciences reviewed EPA's draft IRIS assessment for formaldehyde and found a number of deficiencies that precluded confidence in the weight-of-evidence conclusions drawn in the IRIS assessment (NRC, 2011). Because many of the deficiencies noted by the NRC committee are also apparent in the ISA ozone assessment, we have summarized them in Appendix A and noted them throughout the text.

2 Controlled Exposure Studies

There are four controlled exposure studies of which we are aware that assess the association between ozone and lung function at exposures below 80 ppb (Adams, 2002, 2006; Schelegle *et al*., 2009; Kim *et al*., 2011). Together, these studies indicate a threshold below 70 ppb at which there are no statistically significant adverse effects associated with ozone, and this is consistent with biological data that support a threshold mechanism of action. Effects at 60 ppb are not adverse, nor do they occur statistically more often than do those associated with filtered air (FA) exposures. Furthermore, any large decrements in lung function in particular individuals at this exposure level cannot be attributed to ozone. In addition, one cannot determine whether there would be a population shift in lung function decrements at 60 ppb ozone exposure based on the small number of individuals in the controlled exposure studies.

After a brief description of the four controlled ozone exposure studies, these issues are discussed in more detail below.

2.1 Controlled Exposure Study Protocols

Adams (2002) investigated lung function effects and respiratory symptoms in 30 healthy, non-smoking young adults (15 of each sex; mean age of 22 years) exposed to ozone for 6.6 hours. He used five experimental protocols with square-wave exposures, in which ozone concentrations were maintained at a constant value throughout the exposure period. Two protocols included exposure to 120 ppb ozone (face-mask or chamber exposure); the other protocols included face-mask exposures at 80 ppb and 40 ppb ozone and a chamber exposure to FA. The subjects performed quasi-continuous exercise (QCE; 50 minutes of continuous exercise followed by 10 minutes of rest each hour) during the exposure period, alternating between a cycle ergometer and a treadmill each hour. A minute ventilation rate (V_E) of ~ 20/L/min/m^2 of body surface area (BSA) was maintained during exercise. After the first three hours of exposure, the subjects were given a brief lunch break during which they were exposed to ozone in chamber (but not face-mask) protocols. The endpoints examined in this study were percent changes in one-second forced expiratory volume (FEV_1) and forced vital capacity (FVC), absolute changes in self-reported total symptom severity (TSS), and pain on deep inspiration (PDI). All endpoints were evaluated before, after, and hourly during the 6.6-hour experimental protocol.

Adams (2006) examined the lung function effects of time-dependent chamber exposures to ozone for 6.6 hours in 30 healthy, non-smoking young adults (15 of each sex; mean age of 23 years). Much of the experimental study design followed the structure of previous studies (Adams, 2002; McDonnell *et al*., 1991; Horstman *et al*., 1990); the concentration-time profile of exposure differed because one of the goals of the study by Adams (2006) was to investigate the differences between triangular and square-wave exposure scenarios. In the triangular exposure scenario, ozone concentrations were increased step-wise each hour for the first four hours, then decreased in the last two hours of exposure to achieve an overall cumulative average equal to a specified level also used as a uniform or square-wave exposure. Adams (2006) used six different experimental protocols: standard square-wave concentration profiles at 0 (FA), 60, and 80 ppb ozone and triangular concentration profiles averaging 40, 60, and 80 ppb ozone. The same endpoints were examined as in the study by Adams (2002), before, after, and at one-hour intervals during exposure. The subjects performed QCE in order to maintain a V_E of 20 L/min/m^2 BSA, a precedent set by earlier studies. A 35-minute lunch break was taken at rest in the chamber after three hours of exposure, with the ozone concentration maintained at that used in the third hour. Characteristics such as age, height, weight, body fat, VO_{2max}, and other baseline pulmonary measures (FEV_1, FVC, FEV_1/FVC) were recorded and summarized, but were not considered in the data analyses as possible confounders.

Schelegle *et al*. (2009) conducted a series of chamber exposure studies on 31 healthy, non-smoking young adults (16 females, 15 males; mean age of 21 years) using 6.6-hour triangular exposure scenarios with mean ambient ozone concentrations of 63, 72, 81, and 88 ppb, as well as FA. The subjects performed QCE as in the studies by Adams (2002, 2006), and a 35-minute lunch break was given after three hours in the chamber at the ozone concentration used during the third hour. Both FEV_1 and TSS were measured at multiple time points up to 6.6 hours of exposure, as well as one hour post-exposure, and, in a subset of 13-17 subjects (depending on concentration), four hours post-exposure.

Kim *et al*. (2011) exposed 59 healthy, non-smoking young adults (32 females, 27 males; mean age of 25 years) to FA and 60 ppb ozone (square-wave) for 6.6 hours under controlled chamber conditions in order to evaluate the change in FEV_1 from baseline, as well as TSS. As in the studies described above, the subjects performed QCE during the exposure period, with a 35-minute lunch break in the chamber. Lung function was measured before beginning the initial QCE period; during the 10-minute rest periods after 3, 4.6, and 5.6 hours; and immediately after the 6.6-hour exposure period, but only the measurements taken before the first hour and after the sixth hour of exposure were used in the analyses.

2.2 A Cross-Study Comparison Indicates a Threshold Below 70 ppb Ozone at Which There are No Significant Adverse Effects

In the ISA, EPA conducted a cross-study analysis of several controlled exposure studies (Adams, 2002, 2003, 2006; Folinsbee *et al*., 1988; Horstman *et al*., 1990; McDonnell *et al*., 1991, 2007), including two of the studies that evaluated exposures less than 80 ppb. On page 6-5, EPA states:

> As illustrated in Figure 6-1, there is a smooth dose-response curve without evidence of a threshold for exposures between 40 and 120 ppb O_3. Taken together, these data indicate that mean FEV_1 is clearly decreased by 6.6-h exposures to 60 ppb O_3 and higher concentrations in subjects performing moderate exercise.

This is inconsistent with several statements by EPA throughout the chapter, such as this statement on page 6-6:

> At 40 ppb, triangular and S-W patterns produced responses similar to FA exposure (Adams, 2002, 2006).

In fact, neither EPA nor CASAC conclude that there are any differences between FEV_1 decrements after exposure to 40 ppb ozone and FA. Even though there is disagreement regarding what occurs at 60 ppb (discussed below), this statement indicates that there is **a threshold of at least 40 ppb** because a threshold, by definition, is an exposure below which no statistically significant effects occur. Thus, EPA should add that 40 ppb is a clear no-effect level for FEV_1 changes to the ISA.

In addition, while EPA acknowledges that effects on FEV_1 aren't statistically significant in individual studies at 40 and 60 ppb (*e.g.*, see pg 6-7), the error bars in Figure 6-1 do not reflect this. EPA should change Figure 6-1 to include error bars that include zero to more accurately reflect the variability (and lack of statistical significance) at 40 and 60 ppb.

The fact that a statistical curve without a threshold can be fit to the data does not in itself provide evidence for a lack of a threshold. As noted above, 40 ppb is a clear no-effect level. Also, the smooth curve in Figure 6-1 is based on the study by McDonnell *et al*. (2007), which is not based on exposures < 80 ppb or any data from the more recent studies by Adams (2002, 2003, 2006), Schelegle *et al*. (2009) and Kim *et al*. (2011); rather, it relies heavily on data from Folinsbee *et al*. (1988), Horstman *et al*.

(1990), and McDonnell *et al*. (1991). In an earlier risk assessment funded by EPA and the Department of Energy, Whitfield *et al*. (1996) stated:

> One particularly bothersome case involved the lung function endpoints for 6.6-h exposures of subjects engaged in moderate exertion (the combined data of the studies by Folinsbee *et al.* 1988, Horstman *et al.* 1990, and McDonnell *et al.* 1991). The observed response rate at 0.12 ppm for the FEV_1 decrement ≥ 15% endpoint was judged to be unreasonable and was not used in the regression.

The fact that FEV_1 data from lower exposures are consistent with the McDonnell *et al.* (2007) curve does not prove the curve is appropriate, as these data would be consistent with other curves, as well. Given the small amount of data on which the curve is based, inaccuracies in the data for the high exposures certainly influenced the low end of the curve. Although not discussed in the ISA, this is consistent with McDonnell *et al.* (2010), who concluded that their FEV_1 model overpredicts low concentration (< 80 ppb) values from the more recent studies:

> When individual E-R data become available for low-*C* exposures, we suspect that a refit of the current model (or one with minor modifications) will result in improved predictions at early time points and low O_3 levels that are typical of ambient exposures.

Finally, if one has information regarding thresholds or the mode of action of an agent, this should inform the statistical curves that are fit. One should not choose a non-threshold curve when the mode of action is clearly threshold in nature and the data support a threshold greater than 40 ppb and below 70 ppb ozone, regardless of the magnitude of effects at 80 ppb and above.

2.3 Antioxidant Defenses Against Ozone Indicate a Threshold Mode of Action

Schelegle *et al*. (2007) assessed the time course of ozone-induced changes in breathing patterns in 97 healthy individuals. They found that if the cumulative inhaled dose of ozone at onset of tachypnea (rapid breathing) was not reached during exposure, FVC and FEV_1 decrements were mild or not present. They suggested that this was consistent with the following mode of action:

> As a result of its high reactivity with organic molecules and its low water solubility, O_3 on inhalation penetrates into the lower respiratory tract where it reacts rapidly with components of ALF [airway lining fluid]. Initially, O_3 reacts with antioxidants such as ascorbic acid, reduced glutathione, and uric acid that are contained in the ALF and act as a defense mechanism against oxidant damage by scavenging free radicals and O_3 (3, 7).

> However, O_3 exposure of sufficient duration and concentration can overwhelm these antioxidants, allowing oxidative damage to occur to airway epithelial cells. (p. 696)

That is, antioxidants react with ozone in a time- and concentration-dependent manner and it is only when these defenses are saturated that ozone can cause adverse effects. Thus, a threshold exists below which these antioxidant defenses are sufficient to protect against adverse effects. This is consistent with each of the controlled exposure studies with exposures less than 80 ppb.

2.4 Mean FEV_1 Decrements at 60 ppb are Not Statistically Significant

EPA claims that the data support an association between 60 ppb ozone and lung function decrements. On pages 6-7 to 6-8 of the ISA, EPA states:

> At 60 ppb, there is information available from 4 separate studies (Adams, 1998, Adams, 2006; Kim *et al.*, In Press; Schelegle *et al.*, 2009). The group mean O_3-induced FEV_1 decrements observed in these studies were 3.6% by Adams (1998), 2.8% (triangular exposure) and 2.9% (S-W exposure) by Adams (2006), 3.5% by Schelegle *et al.* (2009), and 1.8% by Kim *et al.* (In Press). Based on data from these three studies, at 60 ppb, the weighted-average group mean O_3-induced FEV_1 decrement (*i.e.*, adjusted for FA responses) is 2.7% (n=150) (Adams, 1998; Adams, 2006; Kim *et al.*, In Press; Schelegle *et al.*, 2009). Although ***not found to be statistically significant in the original studies***, these group mean changes in FEV_1 at 60 ppb are consistent between studies, *i.e.*, none observed an average improvement in lung function with following a 6.6-h exposure to 60 ppb O_3. [emphasis added]

Because the group mean changes are not statistically significant (*i.e.*, the findings are consistent with there being no effect), one must conclude that the association between 60 ppb ozone and FEV_1 decrements cannot be scientifically interpreted as causal. Despite this, much has been made of the results of the Adams (2006) study. It has been argued recently by some CASAC members that Adams (2006) only found statistically non-significant results because the statistical test, specifically the Scheffe *post hoc* test, was not sufficiently powerful to detect the effect. The Scheffe test used by Adams (2006) is a commonly used statistical test to compare mean values that minimizes false positives, but may be more likely to produce false negatives. Some CASAC members suggested that this test is overly conservative and had other statistical tests been used, results would have been statistically significant (Allen *et al.*, 2011).

Brown (2007) and Brown *et al*. (2008) claim to address this by analyzing only the 6.6-hour response at 60 ppb ozone *vs*. FA in the Adams (2006) study. This approach excluded all pulmonary function data at other interim hourly time points and exposure levels within the 6.6-hour exposure pattern. This approach is also at variance with those of other research groups that have performed prolonged ozone exposures and published their results in the scientific literature prior to the Brown reanalysis, including those by researchers at the University of Rochester (Torres *et al*., 1997), the University of Toronto (Liu *et al*., 1999), the University of California, Los Angeles (Gong *et al*., 1998), and EPA (Gong *et al*., 2004). In a presentation to CASAC, Professor William Adams expressed concerns with EPA's reanalysis of selected data from his study and its conclusions, which are very different from those in the published paper (Adams, 2007). While the approach used by Brown (2007) and Brown *et al*. (2008) produced statistically significant results, this can be attributed to the majority of the data being selectively omitted from the analysis.

Dr. Mark Nicolich conducted an analysis of the full data set from Adams using a mixed model analysis of variance and Dunnett's *post hoc* test instead of the Scheffe test (Deason, 2007). This re-analysis, using a technique that is less likely to produce false negatives, was consistent with the original finding by Adams (2006) that there was no statistically significant decrement in group FEV_1 after exposure to 60 ppb ozone *versus* filtered air after 6.6 hours of exercise. In addition, Lefohn *et al*. (2010) re-analyzed five controlled ozone exposure studies, including those by Adams (2006) and Schelegle *et al*. (2009), using two-factor ANOVA and evaluating statistical significance using the Tukey's studentized range approach to account for multiple comparisons for least square means. Although they did not subtract the FA FEV_1 from the ozone-treatment responses, their methodology can still be considered conservative in that it minimizes Type II errors. They did not find any statistically significant changes in FEV_1 at any measurement time associated with the 40 ppb and 60 ppb profiles. In addition, they found that in four out of five studies, exposures to FA substantially improved FEV_1 response over the exposure period, which means that analyses using FA controls likely overestimated FEV_1 changes.

EPA has given no scientifically acceptable justification for relying on the Brown statistical analyses over the original analyses conducted by the authors or those of Nicolich or Lefohn *et al*. (2010). While each statistical method has strengths and limitations, several scientifically accepted statistical methods indicate that there is no statistically significant association between exposure to 60 ppb ozone and lung function decrements. We recommend that in the ISA, EPA recognize and give equal or greater weight to analyses using these methods.

2.5 Controlled Human Exposure Studies were Not Designed to Assess Individual Data

All of the controlled human exposure studies were designed to assess differences on the group mean level, not changes in a specific individual. Because of this, one cannot determine whether effects in certain individuals, even if quite large, are in fact representative of any generalized effect of ozone exposure.

If one were interested in determining whether changes in lung function in a given individual were due to ozone, one would have to design a study that has several repeat measurements for each individual, and perform a scientifically acceptable statistical test on the data for each individual. Otherwise one cannot know whether individual changes are due to ozone, chance, or some other factor (Pagano and Gauvreau, 2000).

The misleading nature of overinterpreting individual findings is illustrated by a review of the data from the Kim *et al*. (2011) and Adams (2006) studies. Kim *et al*. (2011) reported three people with FEV_1 decrements >10% at 60 ppb. Two of these people also underwent 80 ppb exposures, but had a lesser FEV_1 decrement at 80 ppb than at 60 ppb. Similarly, in the study by Adams (2006), one of two individuals with the square-wave exposures and one of two with triangular exposures also had lower FEV_1 decrements after exposure to 80 ppb than to 60 ppb. Because each of these individuals only had one measurement at 60 ppb and 80 ppb, respectively, at the end of the 6.6-hour exposure protocol (when the statistical test was conducted), one cannot determine, ***based on the data,*** whether or not 80 ppb is actually healthier, the same, or worse than 60 ppb in each of these individuals. Because it is biologically implausible that 80 ppb is generally healthier than 60 ppb, and in fact EPA and CASAC seem to conclude that 80 ppb less healthy than 60 ppb, the anomalous data from these individuals suggests that the individual data from these studies are not a reliable measure of general population response.

In an evaluation of the Adams (2006) and four other controlled exposure studies, Lefohn *et al*. (2010) stated, "Additional common findings, based on our reanalysis, among healthy exercising young adults included (i) high intraindividual variability in subject response within exposure profiles; (ii) inconsistent individual FEV_1 response patterns across exposure profiles; [and] (iii) FA exposure FEV_1 changes up to ± 5% in some subjects…" This indicates that any FEV_1 decrement (or increment) should

have an error bar of at least 5% and that a large proportion of most measured FEV_1 responses must be attributable to factors other than ozone in these studies.

On pages 6-8 to 6-10 of the ISA, EPA presents a discussion on inter-subject variability in responses of healthy subjects. To facilitate this discussion, EPA presents "frequency distributions" of FEV_1 decrements from the study by Schelegle *et al.* (2009) in Figure 6-2. In the discussion and distribution charts, EPA fails to recognize the variation attributable to experimental factors; rather, EPA assigns all variability to individual response differences. We strongly recommend that EPA include error bars on the distribution charts, and include in the discussion that at low ozone exposure concentrations (*e.g.*, 60 ppb), the magnitude of the experimental error and daily variation in FEV_1 is significant compared to the small changes possibly due to test exposure.

EPA considers an FEV_1 decrement > 10% to be adverse. The controlled human exposure studies are informative regarding the exposure level, on average, at which group mean FEV_1 decrements differ from those in FA. They are not informative regarding whether responses in any particular individual are fully caused by ozone, however, even if measured FEV_1 decrements are > 10%. Studies that were not designed to determine whether effects in individual subjects are attributable to ozone should not be reanalyzed or interpreted as such in the ISA.

2.6 Mean FEV_1 Decrements at 60 ppb are Not Adverse

Section 109 of the Clean Air Act directs the EPA Administrator to set and revise a primary NAAQS "to protect against adverse health effects" of criteria pollutants. There is no indication that the reported FEV_1 decrements at 60 ppb in the controlled human exposure studies are adverse.

Regarding what constitutes an adverse effect on pulmonary function, we recommend that EPA add the text from the American Thoracic Society (ATS) publication that states:

> The committee recommends that a small, transient loss of lung function, by itself should not automatically be designated as adverse. In drawing the distinction between adverse and nonadverse reversible effects, this committee recommended that reversible loss of lung function in combination with the presence of symptoms should be considered adverse. (ATS, 2000)

EPA should also note that the European Respiratory Society (ERS) suggests that only short-term changes in FEV_1 exceeding 12% "may be clinically important," and that changes in FEV_1 measurements should exceed 5% to overcome the intra-day variability of FEV_1 in normal subjects (Pellegrino *et al*., 2005).

EPA (2007) has concluded previously that a moderate decrement in lung function (defined as a decrease in FEV_1 between 10% and 20%) *and/or* respiratory symptoms are considered to be adverse. We recommend that EPA clearly indicate that their criteria for adversity differs from that of the ATS, as the ATS recommendation does not consider small changes in lung function alone to be adverse. Also, we note that ATS states that EPA's criteria of 10% has not been validated for acceptability or against other measures (ATS, 2000). We recommend that EPA clarify in the ISA that the 10% criteria for adversity is EPA's own internal criteria and is not based on ATS guidelines because ATS has not provided specific numeric criteria for changes in FEV_1 or other measures of lung function.

Regardless, the FEV_1 decrements reported at 60 ppb ranged from 1.7-3.5% and were not accompanied by an increase in respiratory symptoms. Although some individuals had larger decrements, as discussed above, these cannot be attributed wholly to ozone. Overall, the lung function effects at 60 ppb ozone in controlled exposure studies are within the range of intra-individual variability in normal subjects and are not considered adverse with respect to broadly recognized clinical guidelines (*e.g.*, ATS and ERS).

2.7 One Cannot Determine a Population Shift in FEV_1 Changes from Controlled Exposure Studies Conducted with Subjects with High Ventilation Rates

ATS (2000) suggested that if the relationship between a risk factor and a disease is causal, the shift in the risk factor distribution of the exposed population, and hence the risk profile of the exposed population, should be considered adverse, even in the absence of the immediate occurrence of frank illness. In order to determine whether this criterion is met, one must rely on studies from which it is appropriate to extrapolate to the general population. While the controlled exposure studies provide a conservative estimate of ozone risks (because the enhanced exercise regimens in these studies compensate for compromised respiratory systems in sensitive populations), it is not appropriate to use these studies to determine at what exposure there will be a population shift in FEV_1 decrements.

The controlled exposure study subjects were not selected to be representative of the general US population. They were young, healthy adults who exercised during the study. Ozone responsiveness is highest in young adults, as it has been demonstrated to be halved in children (McDonnell *et al*., 1985) and, after young adulthood, to decrease with increasing age (McDonnell *et al*., 1995, 2007). Lung function measurements in these individuals are not applicable to asthmatics or other sensitive people who are not able to achieve the total cumulative absorbed ozone doses in the controlled exposure studies because of their inability to exercise strenuously for long durations. For example, while the nominal ambient ozone concentration in the Adams (2006) study at issue is 60 ppb ozone, the subjects were exercising vigorously (V_e = ~20 L/min/m^2 BSA) over 6.6 hours, so the absorbed dose of ozone was substantially higher than it would have been were the subjects at rest, and also likely higher than it would be in average workers who conduct manual labor (see Linn *et al*., 1993). Indeed, the absorbed dose can probably be safely said to be much higher than that of sensitive populations exerting even moderate physical effort. Because this is the case, Adams (2006) and other studies conducted with high ventilation rates have a bias toward overestimating the effect on lung function response, providing a margin of safety for sensitive individuals in the population.

In addition, the number of subjects in the controlled exposure studies is far too small to estimate the proportions of individuals that would be affected in the general population. In this vein, it is not appropriate to use frequency distributions (as EPA did in Figure 6-2 on pg 6-10) to assess population shifts in FEV_1 decrements. This is because the proportion of study subjects with FEV_1 decrements ≥10% is too uncertain to determine the proportion that would occur in the general population. Furthermore, these proportions are based on one measurement per individual and they do not take into account intra-individual variability. Given that intra-individual variability is on the order of 5%, it is highly likely that some individual responses are misclassified.

It has been suggested that small changes in FEV_1 that are claimed to be statistically significant but are not clinically relevant are indicative of an adverse effect of ozone, presumably because these small changes indicate a population shift in the risk profile of the exposed population. The controlled exposure studies are useful in that they provide a conservative estimate of ozone risks, but they are not designed to determine at what exposure a population shift in effects occurs.

If EPA elects to continue using the results of chamber studies with a small number of subjects to make broad conclusions about large populations, we recommend that the ISA include a clarification that this approach is at variance with ATS criteria. The ATS (2000) guidelines clearly draw a distinction

between small changes in lung function alone in individuals, as derived from clinical studies (which ATS does not consider to be adverse), and small but significant reductions in the lung function of a population, as derived from observational epidemiology studies (which ATS considers to be clinically significant).

3 Epidemiology Studies

The epidemiology evidence for health effects associated with both short-term and long-term ozone exposure does not support causality as determined in the ISA. For short-term ozone exposure, EPA concludes that a causal relationship exists for respiratory morbidity; there is likely to be a causal relationship with mortality; and the evidence is suggestive of a causal relationship for CV morbidity (US EPA, 2011). For long-term ozone exposure, EPA concludes that there is likely to be a causal relationship with respiratory morbidity, and the evidence is suggestive of a causal relationship for mortality (US EPA, 2011). Below, we describe the reasons why the evidence for these health outcomes does not support these causal determinations, as well as the issues with the interpretation of the evidence by EPA in the ISA.

3.1 Short-Term Respiratory Morbidity

The 2006 AQCD evaluated acute ozone-related respiratory morbidity and stated that the studies included provide "clear evidence of causality for the association observed between acute ozone exposure and relatively small, but statistically significant declines in lung function" (US EPA, 2006). In the ISA, EPA states that the results of recent short-term studies are consistent with the findings in the 2006 AQCD, particularly for respiratory-related hospital admissions and ED visits, and that the collective body of epidemiology evidence supports associations with decrements in lung function. The results of recent studies that examined associations between short-term ambient ozone exposure and acute respiratory morbidity have been mixed across all health endpoints, however, and EPA's approach to assessing the evidence is inappropriate and does not consider many of the limitations and uncertainties associated with the underlying epidemiology studies, as described below.

3.1.1 EPA Frequently Misinterprets or Discounts Evidence from Studies Reporting Null Results

As noted in our pre-draft ISA comments (Goodman and Prueitt, 2010), EPA's assessment of the evidence results in a higher number of studies being deemed as providing "positive" evidence than what the data actually indicate. In the ISA, EPA consistently reports any risk coefficient above null as positive evidence, whether or not the result is statistically significant. In addition, null results are often discounted by the inclusion of qualifiers, such as the limitations noted by the study authors, following the summary of the null findings. These qualifiers imply that results would likely have been positive if not for these

limitations. By contrast, the qualifiers rarely appear after the results of studies reporting positive risk estimates (or those that are incorrectly interpreted as positive if above the null but not statistically significant), providing a false perception that most of the "reliable" evidence for effects is positive.[2]

Several examples of the use of qualifiers to discount null results are found throughout the ISA. EPA discounts the null results of the study of respiratory symptoms by Schildcrout *et al*. (2006), noting that the study may have lacked sufficient power to detect ozone-related effects because the authors restricted their analyses to the summer months. EPA discounts the study by Girardot *et al*. (2006) that reported no clear association between ozone and pulmonary function changes in hikers by emphasizing that some subjects failed to provide at least two acceptable spirometry tests. Elsewhere in the ISA, EPA does not acknowledge the high frequency of unreliable values in studies using self-reported peak expiratory flow rate (PEFR) measurements.

By relying on results that are not statistically significant and focusing on selected results while ignoring others, EPA classifies studies reporting little or no association between ozone and respiratory effects as "positive" studies that contribute to their causality conclusion. For example, EPA claims that the results of the study by O'Connor *et al.* (2008) support its conclusion that current levels of ozone cause lung function decrements in asthmatic children. EPA states that using year-round data, O'Connor *et al*. (2008) reported decreases of 0.41% and 0.22% in FEV_1 and PEFR, respectively, per 20 ppb increase in 24-hour average ozone concentrations (for the average of lags 1-5). EPA does not note in the ISA that neither of these results were statistically significant or that the magnitude of the response for ozone was small compared to that reported for other pollutants. EPA also fails to note that O'Connor *et al*. (2008) reported no association between ozone and asthma symptoms or school absences, although in a later section of the ISA, the reported association with asthma symptoms (wheeze) is incorrectly stated as being positive. Finally, EPA fails to note that O'Connor *et al*. (2008) concluded that the associations for ozone in their study were very weak compared to those in their previous study (Mortimer *et al*., 2002). EPA placed very high reliance on the study by Mortimer *et al*. (2002) in the 2008 ozone NAAQS review and subsequent reconsideration, even though it was suggested in the 2008 NO_2 ISA (US EPA, 2008) that the study was unreliable because of the use of self-reported PEFR data.

[2] This is consistent with observations by the NRC (2011) committee, which noted that EPA appeared to give equal weight to all publications regardless of their limitations in the formaldehyde assessment. This issue is apparent in the ISA, as all studies with positive results appear to be given equal weight, without discussion of their limitations, and studies with negative result are given less weight.

3.1.2 EPA Does Not Adequately Consider Factors that Can Bias Study Results

There are many factors that could have biased the results of epidemiology studies of short-term exposure to air pollutants, such as exposure or outcome measurement error and the choice of lag period. EPA does not adequately consider these factors in the ISA, and in some cases, makes statements about these factors with an *a priori* assumption that the ozone-respiratory effect relationship is causal. For example, EPA states the limitations of symptom reports in studies of respiratory symptoms and medication use, and then states that these limitations are sources of random measurement error that can bias effect estimates toward the null or increase the uncertainty around effect estimates. EPA should indicate that these limitations can also bias effect estimates *away* from the null.

EPA does not consider the inadequacy of using ambient ozone measurements from central monitors as a surrogate for personal exposure to ambient ozone in its assessment of the epidemiology data. Many studies confirm the very low correlation between ambient and personal ozone exposure (*e.g.*, Sarnat *et al*., 2001, 2005, 2006), leading to exposure measurement error which can bias the results of an epidemiology analysis in either direction. This exposure measurement error has profound implications for assessing both causality and ozone concentration-response considerations (Rhomberg *et al*., 2011a). It is a considerable source of uncertainty that should be discussed just prior to the review of epidemiology studies in the ISA and should be considered when weighing the evidence from studies that used central monitors for exposure data.

EPA does not adequately take into account the biological plausibility of lag times for effects on pulmonary function. The extensive human clinical and mechanistic data on ozone indicate that the respiratory effects of ozone occur soon after exposure; therefore, findings for 0-1 day lags or for cumulative ozone exposure over a few days should carry more weight in EPA's assessment of the evidence. Often, studies that examined multiple lag times reported positive associations only from one or a few biologically implausible lag times, but not from plausible lag times, which weakens the evidence for causality. This is not apparent in the ISA, however, as EPA often emphasizes results from lag models that produce positive associations rather than results from biologically plausible lag models, which are often null. For example, EPA places high reliance on the results of the study by Mortimer *et al*. (2002), who evaluated the effects of air pollution on changes in PEFR and respiratory symptoms in asthmatic children. The results for the more biologically plausible lags of 1 and 2 days suggested little risk; rather, the "positive" findings in this study were driven by the results at biologically implausible lags of 3-5 days.

In addition, if positive associations are reported with different lag times across different studies, this does not demonstrate consistent evidence for an association, and this should be considered in the ISA.

3.1.3 EPA Does Not Appropriately Weigh the Evidence

In the ISA, EPA does not appear to weigh the evidence for the various respiratory outcomes examined, as well as for the determination of the causal relationship between short-term ozone exposure and respiratory morbidity. Instead, EPA provides summarizing statements that emphasize only the few positive associations that may have been reported for a given outcome. For example, EPA's summary of the epidemiology data for pulmonary inflammation and oxidative stress states that many recent studies reported positive associations, yet throughout that section, EPA states that the results are very mixed and inconsistent. Another example can be seen in the discussion of respiratory symptoms. EPA states there is a "strong" body of evidence demonstrating associations between ozone and respiratory symptoms among asthmatic children and adults, yet almost all are null in both single- and multi-pollutant models. EPA does not provide a discussion of the strengths and weaknesses of the studies to support the overemphasis of studies with "positive" results.[3]

3.1.4 Conclusions for Short-term Respiratory Morbidity

Recent studies that examined associations between short-term ambient ozone exposure and respiratory morbidity have reported mixed results across all health endpoints. EPA's approach to assessing this evidence in the ISA is inappropriate and does not consider many of the limitations and uncertainties associated with the underlying studies. EPA focuses on the positive associations reported in these studies, whether they are statistically significant or not, and often discounts null results. EPA does not adequately consider the factors that can bias study results, such as exposure or outcome measurement error and the choice of lag period. EPA also does not appear to appropriately weigh the evidence for causality, and provides summarizing statements that emphasize only positive associations. inappropriate scientifically valid assessment of the evidence for short-term respiratory morbidity does not support a causal association.

[3] This was a major issue discussed by the NRC (2011) committee in its review of EPA's formaldehyde assessment, as well.

3.2 Short-Term Cardiac Morbidity

The 2006 AQCD stated that the available evidence was highly suggestive that ozone contributes to CV-related morbidity, but was limited and did not fully substantiate links between ambient ozone and adverse CV outcomes. In the ISA, EPA states that the epidemiologic evidence of CV morbidity is still limited and lacks coherence among specific endpoints, and that recent animal toxicology studies provide stronger evidence for CV morbidity than epidemiology studies. EPA states that based on the relatively strong body of toxicology evidence, as well as consistent evidence of an association between ozone and CV mortality, but weak coherence and biological plausibility for ozone-induced CV morbidity, the generally limited body of evidence is suggestive of a causal relationship between short-term ozone exposure and CV morbidity. While it is true that the available epidemiology evidence in inconsistent and does not support an association between short-term ozone exposure and CV morbidity, there is not consistent evidence of an association between ozone and CV mortality, as EPA states (discussed in more detail below).

To justify its conclusion for suggestive evidence, EPA uses findings from recent experimental animal studies. The key animal studies that EPA relies on were conducted at very high exposure levels (*e.g.*, 584 ppb), however, which have little relevance to ambient human ozone exposures. EPA considers an increase in heart rate variability (HRV) as the key indicator of effect in the animal studies, but the epidemiology evidence regarding HRV is mixed and does not corroborate the animal data. The few epidemiology studies reporting an association with HRV found a *decrease* in this parameter, rather than an increase. In addition, EPA correctly notes that almost all of the recent studies that have evaluated the association between ozone and CV hospital admissions report no association and that the overall evidence is inconclusive.

Overall, the evidence for acute CV morbidity and mortality is not sufficient to conclude there is a "suggestive" causal relationship between short-term ozone exposure and CV morbidity. There is no definition of consistency of findings among studies and no apparent criteria for weighing the evidence, which leads to the appearance that the conclusions made by EPA regarding causation are based on a subjective view of the overall data.[4]

[4] The appearance of basing causal conclusions on a subjective view of the data was discussed by the NRC (2011) committee in its review of EPA's formaldehyde assessment, as well.

3.3 Short-Term Mortality

In the 2006 AQCD, EPA stated that there was strong evidence for associations between short-term ozone exposure and all-cause mortality and that consistently positive associations were reported for ozone-related CV mortality. The few positive, statistically significant risk estimates reported in mortality studies in the 2006 AQCD were very weak and susceptible to confounding and bias, however. EPA acknowledged that multiple uncertainties remained in the assessment of the ozone-mortality relationship (US EPA, 2006), and studies that address these uncertainties are the main focus of this section of the ISA. Some of these studies reported positive associations (*e.g.*, Bell and Dominici, 2008; Zanobetti and Schwartz, 2008), whereas other studies that reported no association between ozone and mortality are not included (Dominici *et al*., 2005; Goldberg *et al*., 2006), even though these studies were mentioned in EPA's provisional assessment of recent ozone data (US EPA, 2009). The omission of studies with null results in the ISA without a clear presentation of the criteria for study inclusion and exclusion indicates a lack of transparency in the assessment.[5]

In the ISA, EPA cites recent studies that focused primarily on three areas of uncertainty: confounding, effect modification (*i.e.*, sources of heterogeneity across cities and the ozone-mortality concentration-response function), and model specification. Overall, the results of these studies indicate that important uncertainties remain for the association between short-term ozone exposure and mortality, and these are not adequately addressed in the ISA. Specifically, the new studies indicate that the ozone-mortality association is significantly confounded by various forms of PM, the currently used ozone time-series models are very sensitive to model specification, and the mortality association varies greatly by region and, thus, is not consistent. In addition, as described above, the evidence for an association between ozone and acute CV morbidity is inadequate, so there is a marked lack of coherence for an association between short-term ozone exposure and CV mortality. Thus, we recommend that EPA change its causality conclusion to reflect the fact that the current evidence is insufficient to establish a causal relationship between short-term ozone exposure and mortality.

[5] This issue was also noted in the review of EPA's formaldehyde assessment by the NRC (2011) committee.

3.3.1 EPA Does Not Adequately Acknowledge that the Relationship Between Short-term Ozone Exposure and Mortality is Significantly Confounded by Other Pollutants

Confounding is a common issue in observational studies of ozone-related health effects. As noted in our pre-draft ISA comments (Goodman and Prueitt, 2010), recent evidence indicates that the ozone-mortality relationship is significantly confounded by PM and sulfate. EPA does not appropriately consider this source of uncertainty in the ISA, however.

As observed in the sections on respiratory and CV morbidity discussed above, EPA often cites null results as positive evidence when discussing the potential confounding of ozone risk estimates by PM or other pollutants. For example, EPA states that for the studies by Bell *et al*. (2007) and Stafoggia *et al*. (2010), risk estimates were "robust" to inclusion of co-pollutants in multi-pollutant models when in fact the results in co-pollutant models were null. When discussing the results of other studies, however, EPA does recognize that various forms of PM are confounding the ozone-mortality relationship. The ISA notes that in the study by Smith *et al*. (2009), inclusion of PM_{10} with ozone in a two-pollutant model resulted in a 22-33% decrease in the ozone-mortality association. Similarly, the ISA notes that Franklin and Schwartz (2008) reported a 31% decrease in the ozone-mortality risk estimate when sulfate PM was included in the model. By concluding that the relationship between short-term ozone exposure and mortality is "likely causal," EPA appears to place little weight on the significant confounding by PM reported in these recent studies. EPA should consider more thoroughly this important source of uncertainty when assessing studies of short-term exposure to ozone and mortality.

3.3.2 EPA Does Not Adequately Consider the Sensitivity of Mortality Risk Estimates to Model Specification

The 2006 AQCD concluded that ozone effects on many different health outcomes are robust to varying model specifications, but this is not generally not the case. The results of many studies, including meta-analyses of ozone mortality time-series studies, have indicated that model selection has a key role in the determination of results. For example, ozone mortality estimates have been shown to be sensitive to the degrees of freedom selected for smoothing long-term trend (HEI, 2003; Ito *et al*., 2005; Katsouyanni *et al*., 2009). In addition, the APHENA study, which included datasets from US, Canadian, and European multi-city studies, indicated that ozone mortality estimates are sensitive to smoothing function type, as well (Katsouyanni *et al*., 2009). In the ISA, EPA only presents the positive associations reported in this study from the use of natural splines because "alternative spline models have been previously shown to

result in similar effect estimates (HEI, 2003)." The APHENA study reported large differences with penalized *vs*. natural splines, however, as results were negative when penalized splines were used and positive when natural splines were used. Although EPA provides a justification for why it did not present the APHENA results from both smoothing functions, this justification does not make sense when the large APHENA study, which is heavily relied upon in the ISA, indicates that there is sensitivity of risk estimates to the type of smoothing function used in the model.

3.3.3 EPA Does Not Adequately Consider the Unexplained Regional Heterogeneity in Ozone-Mortality Risk Estimates

EPA notes that recent multi-city studies, including those by Bell and Dominici (2008) and Smith *et al.* (2009), reported marked and unexplained regional heterogeneity in ozone mortality coefficients. This heterogeneity limits the usefulness of a nationwide combined risk estimate for mortality associated with short-term ozone exposure. In Chapter 2 of the ISA, EPA concludes that the relationship between short-term ozone exposure and mortality is "likely causal," placing little weight on these regional differences. EPA does not explain how the "likely causal" determination would apply to a region of the country where no association has been observed or how this heterogeneity would be factored into a human health risk assessment. Presumably, EPA would apply the "likely causal" conclusion throughout the US and assume that mortality is increased in all regions, even though the actual data indicate otherwise.

The heterogeneity of risk estimates across regions also demonstrates that EPA's conclusion in the ISA that recent multi-city studies reported "consistent" positive associations between short-term ozone exposure and mortality is incorrect. EPA should consider this heterogeneity as an important uncertainty when evaluating the effects of short-term ozone exposure on mortality.

3.3.4 EPA Inappropriately Interprets Mixed Evidence for an Ozone Concentration-Response Threshold as No Evidence to Support a Threshold

In the ISA, EPA describes several recent studies that evaluated the ozone-mortality concentration-response relationship (Bell *et al.*, 2006; Xia and Tong, 2006; Stylianou and Nicolich, 2009; Katsouyanni *et al.*, 2009). EPA concludes that these four studies do not provide evidence that supports a threshold for ozone-mortality associations. This statement is not accurate, however, because Stylianou and Nicolich (2009) and Xia and Tong (2006) reported the existence of thresholds. EPA does not present

the findings of another recent study by Smith *et al*. (2009) that provides evidence of a threshold. Using a reverse subset approach, Smith *et al.* (2009) evaluated the ozone-mortality association above various cutoffs in the range of 15-60 ppb. The authors reported different slopes within the three brackets they evaluated (0-40, 40-60, and 60-80 ppb), indicating a non-linear concentration-response relationship.

There are several factors that limit the ability of studies to assess ozone concentration-response thresholds. It is well known that exposure measurement error can bias regression results, which tends to flatten and apparently linearize a steeper and perhaps even threshold-bearing curve, producing a false linear result (Rhomberg *et al*., 2011a,b). In the ISA, EPA does not present the conclusions from the study by Brauer *et al.* (2002), in which exposure misclassification and threshold concentrations in time-series analyses of air pollution health effects were evaluated. For pollutants such as ozone, which exhibit a very low correlation between ambient and personal exposure, Brauer *et al* (2002) reported that it is not possible to determine whether or not a threshold exists. The authors reported that the use of poorly correlated ambient air measurements as a surrogate for personal exposure obscures the ability to detect thresholds. Another issue is that heterogeneity across cities makes it difficult to identify a threshold (Rhomberg *et al*., 2011b), and EPA acknowledges this throughout the concentration-response section of the ISA.

The conclusion in the ISA that there is no evidence to support a threshold for short-term ozone exposure and mortality indicates that EPA is overstating the limited ability to detect a threshold as a lack of evidence to support one. We recommend that EPA revise its conclusions on thresholds to reflect that the existing studies evaluating the ozone-mortality concentration-response relationship reported mixed results, that methodological issues limited the ability of these studies to discern thresholds, and that both the controlled human exposure studies as well as proposed modes of action for respiratory morbidity support the existence of a threshold.

3.3.5 EPA Does Not Provide Information Regarding How It Weighs the Evidence for Causation

EPA does not appear to appropriately weigh the evidence regarding associations between short-term ozone exposure and cause-specific mortality. The ISA states that new multi-city studies show evidence of associations between ozone exposure and both CV- and respiratory-specific mortality, and that respiratory mortality associations are strengthened in the summer season. This is not true, however, as in all-year analyses, *all* risk estimates for respiratory mortality were null, as were the majority of estimates for CV mortality. The fact that all respiratory mortality risk estimates were null does not

provide any evidence of an association. In the summer, mixed results were reported across studies for both respiratory and CV mortality. EPA does not provide a discussion of the strengths and weaknesses of these studies to support its conclusion that they provide evidence of an association even though many of these studies reported null results.

3.3.6 Conclusions for Short-Term Mortality

There are multiple uncertainties in the assessment of the ozone-mortality relationship, such as confounding, model specification, the ozone-mortality concentration-response function, and regional heterogeneity. The studies reviewed in the ISA provide further evidence that these sources of uncertainty are evident, but EPA does not adequately address them in the ISA. In addition, EPA does not appropriately weigh the evidence for cause-specific mortality, which indicates a large number of null results.

3.4 Long-Term Respiratory Morbidity

The 2006 AQCD concluded that the evidence was inconclusive regarding the association between long-term ozone exposure and respiratory morbidity. In the ISA, EPA states that recent studies report consistent associations with new-onset asthma related to genotype, and provide evidence for associations with respiratory symptoms in asthmatics and for first asthma hospitalization. Recent evidence does not provide support to strengthen this association to a "likely to be causal relationship" as concluded in the ISA, however.

The long-term studies of respiratory morbidity suffer from many of the same limitations as the short-term studies, such as exposure measurement error from the use of central monitors and confounding by other pollutants. As in the ISA sections on short-term health effects, these limitations are not adequately addressed in this section of the ISA.[6]

A few studies are cited in which associations are reported between ozone exposure and new-onset asthma or respiratory symptoms in children with certain gene variants. These data are limited, however, in that essentially only one study of each specific gene variant is discussed. In addition, while some of these studies report reduced risks for these outcomes in children with a specific, "protective" variant in

[6] The NRC (2011) committee noted that this was also an issue in EPA's assessment of formaldehyde.

low ozone communities compared to high ozone communities, it is not demonstrated that children with the alternative variant are actually at increased risk for these outcomes with ozone exposure. For example, the study by Islam *et al*. (2008) reports an interaction *p*-value of 0.003 for the effect of gene variation on the risk of new-onset asthma in children by community-specific ozone level (continuous). This interaction is inconsistent with the fact that Islam *et al*. (2008) also reported null risk estimates for all children (with or without the protective gene variant) in communities with higher ozone concentrations, suggesting that none of the children in the study are at increased risk of new-onset asthma with increasing ozone exposure.

Overall, the determination of the relationship between long-term ozone exposure and respiratory morbidity as "likely to be causal" is questionable. The recent evidence cited for this causal determination mainly comes from studies of asthma-related outcomes in children with specific gene variants, but these studies do not demonstrate any consistent associations between ozone exposure and these outcomes. It is unclear how this evidence is more compelling than that of other health outcomes for which the evidence was determined by EPA to only be "suggestive" of a causal relationship.

3.5 Long-Term Mortality

The 2006 AQCD concluded that an insufficient amount of evidence existed to suggest a causal relationship between long-term ozone exposure and mortality. In the ISA, EPA cites multiple ozone-mortality studies that reported null results, and cites two studies conducted since the last review (one of which reported null results) as new evidence. EPA also states that results of short- and long-term respiratory morbidity studies provide biological plausibility for mortality due to respiratory disease, but then concludes that the collective evidence is suggestive of a causal relationship between long-term ozone exposure and *all-cause* mortality, not just respiratory mortality. Multiple studies reporting no association between long-term ozone exposure and mortality, with one study providing weak evidence for respiratory-specific mortality (as discussed further below), do not provide suggestive evidence for a causal relationship, however.

Several large studies that examined long-term ozone exposure and respiratory or cardio-pulmonary mortality have not reported positive associations. No associations were reported for cardio-pulmonary mortality in the Harvard Six Cities Study by Dockery *et al*. (1993), the American Cancer Society (ACS) study by Pope *et al*. (2002), or the Adventist Health Study of Smog (AHSMOG) by Abbey *et al*. (1999). Abbey *et al*. (1999) also reported no association between long-term ozone exposure and

non-malignant respiratory mortality. A recent study by Wang *et al*. (2009) examined cardio-respiratory mortality in Australia and found that long-term exposure to SO_2, but not ozone, was associated with this endpoint. In the ISA, EPA describes these studies and acknowledges that the available data regarding long-term ambient ozone exposure and either respiratory or cardio-pulmonary mortality, with the exception of one study by Jerrett *et al*. (2009), show no association.

In the ISA, EPA focuses on the follow-up analysis of the ACS cohort by Jerrett *et al*. (2009) in its assessment of the association between long-term ozone exposure and respiratory mortality. No other studies of the ACS cohort have reported associations with cardio-pulmonary mortality that were robust to inclusion of co-pollutants, and the Jerrett *et al*. (2009) study does not provide clear evidence of an association for several reasons. Jerrett *et al*. (2009) reported a weakly positive risk estimate in a multi-pollutant model with $PM_{2.5}$. The authors did not adequately control for the potential confounding effects of co-pollutants, however. Although the study examined ozone air concentrations from 1977 to 2000, only two years of data on $PM_{2.5}$ (1999-2000) were considered because of limited availability of data prior to 1999. Because the levels of ozone and $PM_{2.5}$ decreased considerably between 1977 and 2000, the analysis of ozone included higher levels observed in the past, whereas the analysis of potential confounding by $PM_{2.5}$ considered the more recent, lower levels observed in 1999 and 2000. Furthermore, the exposure metric for ozone focused on daily maximum hourly levels in the warm seasons, whereas for $PM_{2.5}$ the annual average concentration was used. Thus, this approach increased the potential to observe an association between ozone and mortality and decreased the potential to observe $PM_{2.5}$ as a confounder of this association. The authors noted this limitation, stating, "Since particulate air pollution has probably decreased in most metropolitan areas during the follow-up interval of our study, it is likely that we have underestimated the effect of $PM_{2.5}$ in our analysis." Another limitation of the study is that confounding by other pollutants, such as SO_2, was not examined. In an earlier study of the ACS cohort, SO_2 demonstrated a stronger association with mortality than $PM_{2.5}$ (Krewski *et al*., 2000). Because of this, as well as the likely underestimation of confounding by $PM_{2.5}$, the study by Jerrett *et al*. (2009) does not demonstrate an association between ozone and respiratory mortality that is independent of other co-pollutants.

Other aspects of the Jerrett *et al*. (2009) study that are not consistent with a positive association between long-term ambient ozone exposure and mortality include a small inverse association between ozone and mortality from CV disease, ischemic heart disease, and all causes combined, as risk estimates for these outcomes were less than one and statistically significant in two-pollutant models with $PM_{2.5}$. The magnitude of these risk estimates was similar to that of the positive risk estimate for respiratory mortality, and it is not biologically plausible that ozone exposure would be protective of mortality; thus, it

is likely that both positive and negative associations of this magnitude, even if they are statistically significant, are not reliable. There is also high regional heterogeneity in risk estimates, as positive associations were only reported in two of the seven regions examined. Because of this high geographic heterogeneity, it was inappropriate for Jerrett *et al*. (2009) to combine data across cities for a US national risk estimate. Finally, socioeconomic data was collected for the ACS study in 1982-1983 but was never updated, so this potential confounder was not fully accounted for in the analysis.

Although several studies have been conducted, only the study by Jerrett *et al*. (2009) has reported an association with long-term respiratory mortality, and this study had many limitations that weakened the evidence for causality. A casual relationship is also not coherent with the overall inconsistent evidence for respiratory morbidity and the almost exclusively null evidence for short-term effects on respiratory mortality, as described above. Together, the evidence for an association between long-term exposure to ozone and mortality is overwhelmingly null, and one weakly positive study with many limitations is not sufficient to suggest a causal relationship.

References

Abbey, DE; Nishino, N; McDonnell, WF; Burchette, RJ; Knutsen, SF; Lawrence Beeson, W; Yang JX. 1999. "Long-term inhalable particles and other air pollutants related to mortality in nonsmokers." *Am. J. Respir. Crit. Care Med*. 159:373-382.

Adams, WC. 2002. "Comparison of chamber and face-mask 6.6-hour exposures to ozone on pulmonary function and symptoms responses." *Inhal. Toxicol.* 14(7):745-764.

Adams, WC. 2003. "Comparison of chamber and face mask 6.6-hour exposure to 0.08 ppm ozone *via* square-wave and triangular profiles on pulmonary responses." *Inhal. Toxicol.* 15:265-281.

Adams, WC. 2006. "Comparison of chamber 6.6-h exposures to 0.04-0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses." *Inhal. Toxicol.* 18(2):127-136.

Adams, WC. 2007. "Public Comment to CASAC Ozone Review Panel Teleconference." University of California, Davis, CA, 2p., March 5.

Allen, G; Balmes, J; Gauderman, J; Henderson, R; Hopke, PK; Sheppar, L; Speizer, F; Suh, H. 2011. "Preliminary Individual Comments on the Ozone Reconsideration." 24p., February 14.

American Thoracic Society (ATS). 2000. "What constitutes an adverse health effect of air pollution?" *Am. J. Respir. Crit. Care Med*. 161:665-673.

Bell, ML; Dominici, F. 2008. "Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities." *Am. J. Epidemiol.* 167(8):986-997.

Bell, ML; Kim, JY; Dominici, F. 2007. "Potential confounding of particulate matter on the short-term association between ozone and mortality in multisite time-series studies." *Environ. Health Perspect.* 115(11):1591-1595.

Bell, ML; Peng, RD; Dominici, F. 2006. "The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations." *Environ. Health Perspect*. 114(4):352-356.

Brauer, M; Brumm, J; Vedal, S; Petkau, AJ. 2002. "Exposure misclassification and threshold concentrations in time series analyses of air pollution health effects." *Risk Anal.* 22(6):1183-1193.

Brown, JS. [US EPA]. 2007. Memo to Ozone NAAQS Review Docket (OAR-2005-0172) re: The effects of ozone on lung function at 0.06 ppm in healthy adults. 8p., June 14.

Brown, JS; Bateson, TF; McDonnell, WF. 2008. "Effects of exposure to 0.06 ppm ozone on FEV1 in humans: A secondary analysis of existing data." *Environ. Health Perspect.* 116(8):1023-1026.

Deason, D. [Exxon Mobil Corp.] 2007. Letter Report to US EPA, Air and Radiation Docket re: Comments on EPA's Proposed "National Ambient Air Quality Standards for Ozone," 72 *Fed Reg*. 37,818 (July 11, 2007). Docket No. EPA-HQ-OAR-2005-0172, 154p., October 9.

Dockery, DW; Pope, CA; Xu, X; Spengler, JD; Ware, JH; Fay, ME; Ferris, BG Jr.; Speizer, FE. 1993. "An association between air pollution and mortality in six U.S. cities." *N. Engl. J. Med*. 329(24):1753-1759.

Dominici, F; McDermott, A; Daniels, M; Zeger, SL;, Samet, JM. 2005. "Revised analyses of the National Morbidity, Mortality, and Air Pollution Study: mortality among residents of 90 cities." *J Toxicol Environ Health A.* 68(13-14):1071-92.

Folinsbee, LJ; McDonnell, WF; Horstman, DH. 1988. "Pulmonary function and symptom responses after 6.6-hour exposure to 0.12 ppm ozone with moderate exercise." *J. Air Waste Manag. Assoc.* 38:28-35.

Franklin, M; Schwartz, J. 2008. "The impact of secondary particles on the association between ambient ozone and mortality." *Environ. Health Perspect*. 116(4):453-458.

Girardot, SP; Ryan, PB; Smith, SM; Davis, WT; Hamilton, CB; Obenour, RA; Renfro, JR; Tromatore, KA; Reed, GD. 2006. "Ozone and PM2.5 exposure and acute pulmonary health effects: a study of hikers in the Great Smoky Mountains National Park." *Environ Health Perspect*. 114(7):1044-52.

Goldberg, MS; Burnett, RT; Yale, JF; Valois, MF; Brook, JR. 2006. "Associations between ambient air pollution and daily mortality among persons with diabetes and cardiovascular disease." *Environ Res*. 100(2):255-67.

Gong, H; Linn, W; Terrel, S; Anderson, K; Clark, K; Sioutas, C; Casio, W; Alexis, N; Devlin, R. 2004. "Exposure of elderly volunteers with and without chronic obstructive pulmonary disease to concentrated ambient fine particulate pollution." *Inhalation Toxicol.* 16:731-744.

Gong, H; Wong, R; Sarma, RJ; Linn, WS; Sullivan, ED; Shamoo, DA; Anderson, KR; Prasad, SB. 1998. "Cardiovascular effects of ozone exposure in human volunteers." *Am. J. Respir. Crit. Care Med.* 158(2):538-546.

Goodman, JE; Prueitt, RL. 2010. "Comments on US EPA's Health Effects Draft Materials for the Ozone Integrated Science Assessment (ISA)." Gradient, Report to American Petroleum Institute, 49p., September 9.

Health Effects Institute (HEI). 2003. "Revised Analyses of Time-Series Studies of Air Pollution and Health." 306p., May.

Horstman, DH; Folinsbee, LJ; Ives, PJ; Abdul-Salaam, S; McDonnell, WF. 1990. "Ozone concentration and pulmonary response relationships for 6.6-hour exposures with five hours of moderate exercise to 0.08, 0.10, and 0.12 ppm." *Am. J. Respir. Crit. Care Med.* 142:1158-1163.

Islam, T; McConnell, R; Gauderman, WJ; Avol, E; Peters, JM; Gilliland, FD. 2008. "Ozone, oxidant defense genes and risk of asthma during adolescence." *Am. J. Respir. Crit. Care Med.* 177:388-395.

Ito, K; De Leon, SF; Lippmann, M. 2005. "Associations between ozone and daily mortality: Analysis and meta-analysis." *Epidemiology* 16(4):446-457.

Jerrett, M; Burnett, RT; Pope, CA; Ito, K; Thurston, G; Krewski, D; Shi, Y; Calle, E; Thun, M. 2009. "Long-term ozone exposure and mortality." *N. Engl. J. Med*. 360(11):1085-1095.

Katsouyanni, K; Samet, JM. 2009. "Air Pollution and Health: A European and North American Approach (APHENA)." Health Effects Institute, 132p., October 29.

Kim, CS; Alexis, NE; Rappold, AG; Kehrl, H; Hazucha, MJ; Lay, JC; Schmitt, MT; Case, M; Devlin, RB; Peden, DB; Diaz-Sanchez, D. 2011. "Lung function and inflammatory responses in healthy young adults exposed to 0.06 ppm ozone for 6.6 hours." *Am. J. Respir. Crit. Care Med.* 183:1215-1221..

Krewski, D; Burnett, RT; Goldberg, MS; Hoover, K; Siemiatycki, J; Jerrett, M; Abrahamowicz, M; White, WH. 2000. "Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, Part II: Sensitivity Analyses." Health Effects Institute, Cambridge, MA, p129-293, July.

Lefohn, AS; Hazucha, MJ; Shadwick, D; Adams, WC. 2010. "An alternative form and level of the human health ozone standard." *Inhal. Toxicol.* 22(12):999-1011.

Linn, WS; Spier, CE; Hackney, JD. 1993. "Activity patterns in ozone-exposed construction workers." *J. Occup. Med. Tox*. 2:1-14.

Liu, L; Leech, JA; Urch, RB; Poon, R; Zimmerman, B; Kubay, JM; Silverman, FS. 1999. "A comparison of biomarkers of ozone exposure in human plasma, nasal lavage and sputum." *Inhal. Toxicol.* 11(8):657-674.

McDonnell, WF; Chapman, RS; Leigh, MW; Strope, GL; Collier, AM. 1985. "Respiratory responses of vigorously exercising children to 0.12 ppm ozone exposure." *Am. Rev. Respir. Dis.* 132:875-879.

McDonnell, WF; Kehrl, HR; Abdul-Salaam, S; Ives, PJ; Folinsbee, LJ; Devlin, RB; O'Neil, JJ; Horstman, DH. 1991. "Respiratory response of humans exposed to low levels of ozone for 6.6 hours." *Arch. Environ. Occup. Health.* 46:145-150.

McDonnell, WF; Stewart, PW; Andreoni, S; Smith, MV. 1995. "Proportion of moderately exercising individuals responding to low-level, multi-hour ozone exposure." *Am. J. Respir. Crit. Care Med.* 152:589 596.

McDonnell, WF; Stewart, PW; Smith, MV. 2007. "The temporal dynamics of ozone-induced FEV_1 changes in humans: An exposure-response model." *Inhal. Toxicol.* 19(6-7):483-494.

McDonnell, WF; Stewart, PW; Smith, MV. 2010. "Prediction of ozone-induced lung function responses in humans." *Inhal. Toxicol.* 22:160-168.

Mortimer, KM; Neas, LM; Dockery, DW; Redline, S; Tager, IB. 2002. "The effect of air pollution on inner-city children with asthma." *Eur. Respir. J.* 19:699-705.

National Research Council (NRC). 2011. "Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde (Prepublication Copy)." National Academies Press, Washington, DC, 155p. Accessed on April 8, 2011 at http://www.nap.edu/catalog/13142.html, April.

O'Connor, GT; Neas, L; Vaughn, B; Kattan, M; Mitchell, H; Crain, EF; Evans, R III; Gruchalla, R; Morgan, W; Stout, J; Adams, GK; Lippman, M. 2008. "Acute respiratory health effects of air pollution on children with asthma in US inner cities." *J. Allergy Clin. Immunol.* 121(5):1133-1139.

Pagano, M; Gauvreau, K. 2000. *Principles of Biostatistics - Second Edition*. Duxbury, Pacific Grove, CA, 525p.

Pellegrino, R; Viegi, G; Brusasco, V; Crapo, RO; Burgos, F; Casaburi, R; Coates, A; van der Grinten, CP; Gustafsson, P; Hankinson, J; Jensen, R; Johnson, DC; MacIntyre, N; McKay, R; Miller, MR; Navajas, D; Pedersen, OF; Wanger, J. 2005. "Interpretative strategies for lung function tests." *Eur. Respir. J.* 26(5):948-968.

Pope, CA; Burnett, RT; Thun, MJ; Calle, EE; Krewski, D; Ito, K; Thurston, GD. 2002. "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution." *JAMA* 287(9):1132-1141.

Rhomberg, LR; Chandalia, JK; Long, CM; Goodman, JE. 2011a. "Measurement error in environmental epidemiology and the shape of exposure-response curves." *Crit. Rev. Toxicol.* (In press.)

Rhomberg, LR; Goodman, JE; Haber, LT; Dourson, M; Andersen, ME; Klaunig, JE; Meek, B; Price, PS; McClellan, RO; Cohen, SM. 2011b. "Linear low-dose extrapolation for noncancer health effects is the exception, not the rule." *Crit. Rev. Toxicol.* 41:1-19.

Sarnat, JA; Brown, KW; Schwartz, J; Coull, BA; Koutrakis, P. 2005. "Ambient gas concentrations and personal particulate matter exposures: Implications for studying the health effects of particles." *Epidemiology* 16:385-395.

Sarnat, SE; Coull, BA; Schwartz, J; Gold, DR; Suh, HH. 2006. "Factors affecting the association between ambient concentrations and personal exposures to particles and gases." *Environ. Health Perspect.* 114:649-654.

Sarnat, JA; Schwartz, J; Catalano, PJ; Suh, HH. 2001. "Gaseous pollutants in particulate matter epidemiology: Confounders or surrogates?" *Environ. Health Perspect.* 109(10):1053-1061.

Schelegle, ES; Morales, CA; Walby, WF; Marion, S; Allen, RP. 2009. "6.6-Hour inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans." *Am. J. Respir. Crit. Care Med.* 180(3):265-272.

Schelegle, ES; Walby, WF; Adams, WC. 2007. "Time course of ozone-induced changes in breathing pattern in healthy exercising humans." *J. Appl. Physiol.* 102:688-697.

Schildcrout, JS; Sheppard, L; Lumley, T; Slaughter, JC; Koenig, JQ; Shapiro, GG. 2006. "Ambient air pollution and asthma exacerbations in children: An eight-city analysis." *Am. J. Epidemiol.* 164(5):505-517.

Smith, RL; Xu, B; Switzer, P. 2009. "Reassessing the relationship between ozone and short-term mortality in U.S. urban communities." *Inhal. Toxicol.* 21(Suppl. 2):37-61.

Stafoggia, M; Forastiere, F; Faustini, A; Biggeri, A; Bisanti, L; Cadum, E; Cernigliaro, A; Mallone, S; Pandolfi, P; Serinelli, M; Tessari, R; Vigotti, MA; Perucci, CA. 2010. "Susceptibility factors to ozone related mortality: A population-based case-crossover analysis." *Am. J. Resp. Crit. Care Med.* 182:376 384.

Stylianou, M; Nicolich, MJ. 2009. "Cumulative effects and threshold levels in air pollution mortality: Data analysis of nine large US cities using the NMMAPS dataset." *Environ. Pollut.* 157(8-9):2216-2223.

Torres, A; Utell, MJ; Morow, PE; Voter, KZ; Whitlin, JC; Cox, C; Looney, RJ; Spers, DM; Tsai, Y; Frampton, MW. 1997. "Airway inflammation in smokers and nonsmokers with varying responsiveness to ozone." *Am. J. Respir. Crit. Care Med.* 156:728-736.

US EPA. 2006. "Air Quality Criteria for Ozone and Related Photochemical Oxidants. Volume I." Office of Research and Development, National Center for Environmental Assessment, EPA 600/R-05/004aF, 821p., February.

US EPA. 2007. "Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information (OAQPS Staff Paper)." Office of Air Quality Planning and Standards, EPA-452/R-07-007, 609p., July.

US EPA. 2008. "Integrated Science Assessment for Oxides of Nitrogen." National Center for Environmental Assessment (NCEA) – RTP Division, EPA/600/R-09/071, 260p., July.

US EPA. 2009. "Provisional Assessment of Recent Studies on Health and Ecological Effects of Ozone Exposure." Office of Research and Development, National Center for Environmental Assessment, EPA/600/R-09/101, 55p., September.

US EPA. 2011. "Integrated Science Assessment for Ozone and Related Photochemical Oxidants (First External Review Draft)." National Center for Environmental Assessment (NCEA), EPA/600/R 10/076A, 996p., March.

Wang, XY; Hu, W; Tong, S. 2009. "Long-term exposure to gaseous air pollutants and cardio-respiratory mortality in Brisbane, Australia." *Geospat. Health* 3(2):257-263.

Whitfield, RG; Biller, WF; Jusko, MJ; Keisler, JM. June 1996. "A Probabilistic Assessment of Health Risks Associated with Short-Term Exposure to Tropospheric Ozone." Report to US EPA, US Dept. of Energy (US DOE) ANL/DIS-3, 235 p. Accessed on April 15, 2011 at http://www.epa.gov/ttnnaaqs/standards/ozone/data/riskrep.pdf.

Xia, Y; Tong, H. 2006. "Cumulative effects of air pollution on public health." *Stat. Med.* 25:3548-3559.

Zanobetti, A; Schwartz, J. 2008. "Mortality displacement in the association of ozone with mortality: An analysis of 48 cities in the United States." *Am. J. Respir. Crit. Care Med.* 177(2):184-189.

Appendix A

Appendix A

Many of the issues associated with EPA's assessment of the evidence for health effects associated with ozone exposure in the ISA have been identified in other EPA assessments. For example, a committee assembled by the NRC recently published its *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde* (NRC, 2011), which notes general issues that other NRC committees have identified in a number of EPA's IRIS assessments over the last decade. In this case, the committee stated that the EPA draft formaldehyde assessment "was not prepared in a transparent, consistent fashion," and the committee criticized EPA's hazard identification assessment, noting many issues related to EPA's approach for weighing the evidence. These issues included a lack of information regarding study inclusion criteria, an inconsistent method for evaluating strengths and weaknesses of studies, and the lack of a clear framework for evaluating the weight of the evidence in establishing causation.

The committee found insufficient documentation of methods and criteria for identifying the epidemiologic evidence to be reviewed, noting that the specific databases for literature identification, search terms, results of searches, and exclusion criteria were not listed. The committee stated that an *a priori* presentation of the study selection criteria is critical for determination of causation. The lack of such criteria is one example of the lack of transparency in the assessment.

The committee also noted that the methods for evaluating the strengths and limitations of the selected studies were not provided in the assessment. It also felt that the descriptions and evaluation of the individual epidemiology studies were not consistent among studies, stating "The characterization of the strengths and weaknesses of the studies varies; some studies receive a fuller treatment, including a more extensive assessment of bias and its consequences for estimating effect measures, and others receive less attention." This inconsistency is further noted by the committee in its statement that "In some cases, there is a tendency to describe the studies ultimately selected for the derivation of the RfC in favorable terms," without sufficient consideration of their weaknesses. The committee felt that although some attention was given to methodological concerns, particularly in studies considered to be more informative, it was given in a nonsystematic fashion, as some key methodological limitations were inconsistently mentioned and insufficiently explored. The committee stated that the lack of evidence for a specified format for evaluating studies led to weight-of-evidence evaluations that appeared to give equal weight to

all publications, with no consideration of study quality or validity of exposure concentration measurements.

The committee noted that the assessment did not have a clearly articulated framework for establishing causation on the basis of the weight and strength of evidence, and that there were inconsistencies in the approach that EPA used for evaluating causation for different health endpoints. The committee stated that the implementation of weight-of-evidence guidelines in the assessment "appears to be subjective and not standardized." The committee found that variable detail was provided in how the weight-of-evidence criteria had been applied, noting that uniformly developed discussions applying the criteria could not be identified at appropriate points in the text. The committee also felt that EPA took an *ad hoc* approach to hazard identification for some outcomes, such as asthma.

The concept of consistency of findings within and among studies was not defined in EPA's assessment, and the committee stated that in some cases "the conclusion of causation appears to be based on a subjective view of the overall data." The committee also noted that some concluding statements regarding causation seemed to over-interpret the results. The committee recommended that clear narratives are needed to provide the rationale for conclusions, and that when concluding statements are made that indicate the results support an association in the face of mixed results (*e.g.*, positive, weak, and null studies for a given endpoint), these need to be accompanied by a thorough discussion of the strengths and weaknesses of the studies.

Overall, the committee noted recurring methodological issues in the assessment and provided many recommendations for revision. The committee stated that an *a priori* presentation of the study selection criteria and a clearly articulated framework for weighing the evidence are critical for any determination of causation. It recommended that all key studies "need to be thoroughly evaluated with standardized approaches that are clearly formulated based on the type of research." It also recommended that the approach for weight-of-evidence evaluation in the assessment be revised. The committee stated, "Strengthened, more integrative, and more transparent discussions of weight of evidence are needed. The discussions would benefit from more rigorous and systematic coverage of the various determinants of weight of evidence, such as consistency." Many of the issues identified in the assessment by the NRC committee are also apparent in EPA's recent assessment of the evidence for health effects associated with ozone exposure, as presented in the ISA.