April 28, 2014

MEMORANDUM

SUBJECT: Response to Comments Regarding the Potential Use of a Threshold Model in Estimating the Mortality Risks from Long-term Exposure to Ozone in the *Health Risk and Exposure Assessment for Ozone, Second External Review Draft*

FROM: Erika Sasser, Acting Director /s/
Health and Environmental Impacts Division (C504-02)
Office of Air Quality Planning and Standards
United States Environmental Protection Agency

TO: Holly Stallworth
Designated Federal Officer
Clean Air Scientific Advisory Committee
EPA Science Advisory Board Staff Office

During the meeting of the CASAC Ozone Panel on March 25-27, 2014, at which the second draft of the Health Risk and Exposure Assessment (HREA) for Ozone was reviewed, public comments were provided to the Panel by Dr. Anne Smith on behalf of the Utility Air Resources Group. Among her comments, Dr. Smith recommended a change in the approach taken in the HREA to estimate respiratory mortality risks associated with long-term exposures to ozone, based on the study by Jerrett et al. (2009). Dr. Smith recommended that the HREA evaluate "the effect of estimating the risks using results of a model that detected an effects threshold within the data." [1] Excerpts from Dr. Smith's written comments are provided in Attachment 1 to this letter.

In response to these comments, the Panel requested that EPA evaluate the threshold analysis provided in the supplementary appendix included with the Jerrett et al (2009) article, and include a sensitivity analysis showing the impacts on estimated mortality risk of using a threshold model. This memo describes EPA's planned response to the comments based on our evaluation of the information provided in the Jerrett et al. (2009) paper and additional materials provided by the authors. EPA is currently working on a threshold sensitivity analysis, but will not have results available prior to the Panel teleconference on May 28th. Rather, EPA will provide the results of the sensitivity analysis in the final HREA.

[1] Smith, Anne. 2014. "Key Issues in the February 2014 Draft of the Health Risk and Exposure Assessment for Ozone – Advance Written Materials for CASAC Meeting". Prepared on behalf of UARG, March 13, 2014. Available at:
http://yosemite.epa.gov/sab/sabproduct.nsf/bf498bd32a1c7fdf85257242006dd6cb/84006d7423b29d9b85257b96004a8381!OpenDocument&Date=2014-03-25

EPA has reviewed the supplementary appendix provided in Jerrett et al. (2009). In doing so, a number of questions arose regarding the correct interpretation of the threshold analyses provided, specifically related to interpretation of the model fits for alternative thresholds and the ability to determine whether models with any specific thresholds were significantly better than models with other thresholds. In addition, we determined that in order to be able to provide an appropriate and complete sensitivity analysis, we needed additional data that was not provided in Table 3S of the supplementary appendix of Jerrett et al. (2009). To improve our understanding of these issues and obtain the needed data, we sent a written request to Dr. Michael Jerrett, lead author of the article. This written request is included as Attachment 2 to this letter.

The response from Dr. Jerrett is included as Attachment 3 to this memo. The Jerrett response highlights several factors:

1) In terms of goodness of fit, long-term health risk models including ozone clearly performed better than models without ozone, indicating the improved predictions of respiratory mortality when ozone is included.
2) The model including a threshold at 56 ppb had the lowest log likelihood value of all models evaluated (i.e., linear models and those having thresholds ranging from 40-60 ppb), and thus provided the best overall fit to the data.
3) It is not clear whether the 56 ppb threshold model is a better predictor of respiratory mortality than when using a linear model for this dataset. Using one statistical test, the model with a threshold at 56 ppb was determined to be statistically superior to the linear model. Using another, more stringent, test, none of the threshold models considered were statistically superior to the linear model.
4) Under the less stringent test, although the threshold model produces a superior prediction than the linear model, there is uncertainty about the specific location of the threshold, if one exists. This is because the confidence intervals on the prediction indicate that a threshold could exist anywhere from 0 to 60 ppb. Considerable caution should be exercised in accepting any specific threshold, particularly when the more stringent statistical test indicates there is no significantly improved prediction.

Implications for the Health Risk and Exposure Assessment: Based on this response, our planned approach is to conduct a sensitivity analysis evaluating potential thresholds in the concentration-response function relating long-term ozone exposures with respiratory mortality. Using the set of coefficients provided by Dr. Jerrett, we will estimate respiratory mortality risks associated with the range of threshold models provided in the table included in Attachment 3. For purposes of analytical and presentation efficiency, we will provide risk estimates for models with thresholds ranging from 40 to 60 ppb in increments of 5 ppb (similar to the categories provided in Table 3S in the Jerrett article), with an additional estimate provided at a threshold of 56 ppb representing the best fitting of the threshold models. We will present these results in additional tables in Chapters 7 (urban case study analyses) and 8 (national mortality risk burden) directly following the presentations of the long-term exposure mortality risk estimates based on application of the no-threshold model with ozone as a linear term (our current core estimates).

In terms of discussion of the models and results in the text in both Chapter 7 and Chapter 8, we will add a fuller discussion of the evidence provided in the Jerrett article regarding the potential existence and location of a threshold in the concentration-response function. We will also discuss the implications of this evidence for the confidence we have in the specific quantitative estimates of long-term exposure mortality risk, with specific indications that the large impact that this uncertainty has on the magnitude of risk causes us to have reduced overall confidence in those quantitative estimates.

Implications for the Policy Assessment: Based in part on estimates of the number of respiratory deaths associated with long-term O_3, the second draft Policy Assessment concluded that a standard with an 8-hour averaging time and a revised level (i.e., from 70 to 60 ppb) could provide appropriate public health protection against longer-term O_3 exposures. We do not expect that the addition of the sensitivity analyses discussed above will alter this fundamental conclusion in the final Policy Assessment. Specifically, while threshold models will likely estimate fewer O_3-associated respiratory deaths than the linear model, we expect that alternative 8-hour standards with levels from 70 to 60 ppb will still reduce estimated risks relative to the current standard.

We appreciate the advice of the Panel, and look forward to your comments during the upcoming Panel teleconference on this proposed approach to addressing potential thresholds in estimating the mortality risks from long-term exposure to ozone in the final HREA and PA. Should you have any questions regarding this memorandum, please contact me (919-541-3889; email sasser.erika@epa.gov) Dr. Bryan Hubbell (919-541-0621; email hubbell.bryan@epa.gov), or Karen Wesson (919-541-3515; email wesson.karen@epa.gov).

Attachment 1: Excerpt of written comments provided by Dr. Smith

A third and far more interesting sensitivity analysis could have and should have been performed using yet other results in Jerrett et al.: the effect of estimating the risks using results of a model that detected an effects threshold within the data. The HREA does not even suggest that a sensitivity analysis might be warranted regarding the threshold, even though that was a specific issue addressed in the paper. In fact, with respect to the question of a potential threshold, the HREA only states that evidence of a threshold is "limited" in the epidemiological literature generally, and does not mention that evidence of one was reported in Jerrett et al.5 In fact, Jerrett et al. find that a model with a threshold of 56 ppb provides a better fit to the data than the no-threshold model (using a 1-P model in both cases). The HREA should have provided analyses showing how very sensitive its long-term risk estimates are to this better-fit model.

Using BenMAP, we find that the alternative threshold model makes an enormous difference to risks estimates. National risks computed using the 56 ppb/1-P threshold model are 95% less than the no-threshold 1-P result, and 97% less than the no-threshold 2-P result that is the core estimate emphasized in the HREA. Figure 1 provides a map showing the dramatic differences in estimates of elevated risk of respiratory mortality across the nation projected by each of these three alternative models. The 2-P and 1-P no-threshold models (Figures 1A and 1B, respectively) imply that over 6% of respiratory deaths were hastened by ozone levels in 2006-2008 across the entire nation. In stark contrast, the threshold model indicates that risk is elevated by more than 6% only in a part of Southern California, and that risk is zero in the majority of the U.S. It is also noteworthy that under the 56 ppb threshold model, 10 of the 12 cities studied in the HREA are projected to have zero risk for long-term respiratory mortality at the current standard of 75 ppb. Jerrett et al. notes that the 56-ppb threshold model is not statistically significantly different from the zero-threshold model, but we note that the threshold model is nevertheless the best fit. In fact, all thresholds tested in that paper from 45 to 57 provided a better fit (i.e., a higher log-likelihood function value) than the zero-threshold model, and the 56 ppb case had the best fit of all. The paper dismisses this best-fit threshold estimate as not statistically significant because its p value is 0.06. While this p value means that the no-threshold model assumption cannot be rejected with 95% confidence, it can be rejected with 94% confidence. When the question of significance is so borderline, it is not reasonable to ignore the implications of the alternative models in one's risk calculations.

Despite the 56 ppb threshold model being a better fit, the HREA has adopted the no-threshold model for its core analysis. The only statistical principle that would support adopting the zero threshold model as the core result rather than a better-fitting threshold model would be if medical professionals held an a priori belief is that no threshold exists. In fact, the opposite is the case: medical professionals widely expect a diminution of risk at lower concentrations, and the surprise has been that air pollution epidemiological studies have not generally identified thresholds. The finding of a non-linearity in the association at lower ozone levels by Jerrett et al. is consistent with a priori expectations, and it should not be dismissed simply because one has “only” 94% confidence that a threshold exists in these data. Nevertheless, even if the confidence level were much lower but the threshold model still provided a better fit, the appropriate action for the risk analyst would be to highlight the huge policy significance of this very fine line among the alternative models. As is shown above, the alternative models reported in Jerrett et al. produce estimates of as-is risk that differ by over a factor of 25, and range qualitatively from a localized and modest risk increase to a universally large risk. This is an abject degree of uncertainty about what this paper’s findings tell us about long-term respiratory risk from ozone in the U.S. As there are no other publications to support any particular conclusion out of this paper, long-term ozone mortality risk estimates based on Jerrett et al. (2009) do not merit attention in the HREA. If they are to be included at all, the above sensitivity across the threshold assumptions should be the core result that is presented.[2]

---

[2] Smith, Anne. 2014. “Key Issues in the February 2014 Draft of the Health Risk and Exposure Assessment for Ozone – Advance Written Materials for CASAC Meeting” Prepared on behalf of UARG, March 13, 2014. Available at http://yosemite.epa.gov/sab/sabproduct.nsf/bf498bd32a1c7fdf85257242006dd6cb/84006d7423b29d9b85257b96004a8381!OpenDocument&Date=2014-03-25

Attachment 2: Written request to Dr. Michael Jerrett

Dr. Jerrett-

Per our previous telephone conversation, we have several questions regarding your 2009 publication “Long-term Ozone Exposure and Mortality”, published in the New England Journal of Medicine. During the recent CASAC meeting reviewing the second draft of the Ozone Health Risk and Exposure Assessment, a public commenter raised the issue of which model from your article should be used to estimate the mortality risks associated with long-term exposures to ozone. In your article, you provide a supplementary appendix which includes the results of a threshold analysis. In the main body of the article, you report the results of models without an assumed threshold. You note in the main body of the article that “There was limited evidence that a threshold model specification improved model fit as compared with a nonthreshold linear model (p=0.06).” In the threshold analysis, in Table 3S, you report the log likelihood values for a number of different models, including the non-threshold model (recognizing that the lowest ozone value in the data was 33 ppb), and a series of models with alternative thresholds ranging from 45 to 65 ppb. In that table, you report that the lowest log likelihood value occurred for the model with an assumed threshold at 56 ppb, while at the same time noting that the confidence interval based on the log-likelihood included the model with a threshold of 0 ppb to the model with threshold of 60 ppb.

Our first question is: If the 56 ppb threshold model has the lowest log likelihood value across all models including the no-threshold model, why was that model not presented in the main body of the article, even if it was not significantly different from the no-threshold model?

Our second question is: Given the small differences between the log likelihood values across all of the thresholds that were evaluated, are there any statistically significant differences between the models with thresholds ranging from 45 ppb to 65 ppb? Is it possible to quantify the probability that a true threshold would lie between 45 and 65? Between zero and 45 or zero and 65? Is it meaningful to consider the 56 ppb threshold model a significantly better model compared to models with no threshold or alternative thresholds listed in Table 3-S? We would like to understand better the uncertainty both in the existence of a potential threshold and the uncertainty in the specific value of a threshold.

Our third question is: Similar to the information provided for the 56 ppb threshold model, can you provide the p-values, beta, and standard error for the ozone coefficient for each of the threshold values evaluated in Table 3S, to enable calculation of the long-term mortality risk associated with the full range of potential threshold values.

Thank you for your attention to these questions and your assistance in obtaining the threshold model estimates.”

Attachment 3: Response from Dr. Jerrett

Explanation and Interpretation of Threshold Model presented in

Jerrett M, Burnett RT, Pope III A, Ito K, Thurston G, Krewski D, Shi Y, Calle E, Thun M Long-term ozone exposure and mortality. *New England Journal of Medicine 2009* 360, 1085-1095.

Drs. Burnett and Jerrett have prepared this expanded discussion of the threshold model presented in the above article at the request of Dr. Bryan Hubbell Group Leader, Risk and Benefits Group, Health and Environmental Impacts Division, Office of Air Quality Policy and Standards, Office of Air and Radiation to supply additional information that can be used for estimating the benefits and related uncertainties of reductions in ozone concentrations in the United States. We emphasize that this additional information and interpretation has not undergone peer review or review by our co-authors.

We also note that much progress has been made in methods for determining the shape of dose-response curves since the publication of our article in 2009; the methods used in that article represented our understanding of the best available techniques at that time. We recognize, however, that other methods, using model averaging or targeted parameter estimation, may supply more definitive results. At this time, we do not have the resources to implement more sophisticated analyses the data. The results, therefore, should be taken with the caveat that methods developed more recently could change the results of this paper.

Supplementary Analyses and Explanation

A threshold modeled variable for ozone, x_T, was included as a predictor in the Cox Proportional Hazard survival model of the from:

$$x_T = \begin{cases} 0,\ldots\ldots, x < T \\ x - T, x \geq T \end{cases}$$

where *T* denotes a threshold concentration in ppb and *x* is the ozone concentration in ppb. A series of models were fit for various values of *T*, ranging from 40 ppb to 60 ppb by single ppb increments. Two sets of model results are presented in the Table below, with and without ecological covariates included in the survival model. The ozone coefficient/ppb and its corresponding standard error are given in addition to the -2*log likelihood value. Two addition models were fit, one without ozone and one including a linear term in ozone. The -2*log likelihood values for various threshold model specifications summarize how well the model fits the data. Lower values represent models with a better fit to the data than those models with higher values.

For both model specifications with and without ecological covariates, a threshold value of 56 ppb minimized the log-likelihood, indicating that this value was the best fit among threshold models examined. Note that only a single city, San Francesco, had an ozone concentration below 40 ppb and thus no threshold models were considered with $T < 40$. Large changes in the log-likelihood were observed after including with ozone as a linear term or any of the threshold models of ozone compared to the model without ozone, suggesting that these forms of model with ozone clearly improved predictive power for respiratory mortality in this dataset. Also note that the log-likelihood did not monotonically change with changing threshold concentrations, indicating some instability in the likelihood function (i.e., non-monotonic). This may be due to the lack of data between some threshold values.

Differences between -2*log-likelihood values among models can be informative as to the model fit. The difference between the -2*log-likelihood values from the linear and threshold models at 56 ppb is 20.88-18.39=2.49 for the model without ecological covariates and 58.93-55.39=3.54 for the models with ecological covariates. Deciding on whether the optimal threshold model (i.e. T=56) is a better predictor of respiratory mortality in this dataset is not clear. For example, using the AIC criterion would suggest that any difference greater than 2 would be sufficient to conclude that the threshold model was a superior predictor since the threshold model consists of a single additional estimated parameter compared to the linear model. However, using the more stringent BIC criterion, one would need to observe a difference of log(9896)=9.2, when 9,896 is the number of respiratory deaths recorded. Here we have used the sample size as the number of events in the survival model. There is also controversy in the literature as to whether one should use the number of subjects when determining the BIC (i.e. log(448850)=13.0). If this method of calculating the BIC was used it would suggest no significant evidence that any threshold model was a superior predictor to the linear model. In sum, although the threshold model produces a superior prediction than the linear model when evaluated by the less stringent AIC indicator, there is uncertainty about the specific location of the threshold, if one exists. The confidence intervals on the prediction indicate that the threshold could exist anywhere from 0-60 ppb. As noted, we did not observe a significant threshold by conventional standards (i.e., $p < 0.05$) with our original analyses. This extended analysis suggests considerable caution should be exercised in accepting any specific threshold, particularly when the more stringent BIC indicates there is no significantly improved prediction.

Table: -2*log likelihood values based on threshold concentration response model for ozone concentrations measured from April to September, 1977-2000 in the ACS cohort with follow-up from 1982 to 2000, adjusted for individual risk factors, baseline hazard function stratified by age (single year groupings), gender, and race, with or without adjustment for the ecological covariates.

Threshold (ppb)	Without Ecological Covariates -2*log-lik (-1444)	Without Ecological Covariates Coefficient (se) (x1000)	With Ecological Covariates -2*log-lik (-1437)	With Ecological Covariates Coefficient (se) (x1000)
No Ozone in Model	72.51	NA	97.72	NA
Linear Model	20.88	2.81 (0.808)	58.93	2.86 (0.942)
40	19.35	3.09 (0.899)	57.41	3.12 (0.960)
41	19.11	3.13 (0.900)	57.18	3.16 (0.963)
42	18.81	3.19 (0.904)	56.94	3.22 (0.968)
43	18.87	3.21 (0.912)	56.87	3.26 (0.978)
44	18.91	3.23 (0.921)	56.80	3.31 (0.989)
45	18.96	3.26 (0.929)	56.73	3.36 (1.000)
46	19.04	3.28 (0.937)	56.69	3.40 (1.010)
47	19.12	3.29 (0.945)	56.65	3.44 (1.202)
48	19.22	3.30 (0.952)	56.63	3.49 (1.030)
49	19.40	3.31 (0.962)	56.66	3.52 (1.050)
50	19.60	3.32 (0.971)	56.72	3.56 (1.060)
51	19.71	3.34 (0.982)	56.70	3.61 (1.070)
52	19.50	3.42 (0.996)	56.49	3.72 (1.090)
53	19.15	3.54 (1.020)	56.17	3.86 (1.120)
54	18.88	3.36 (1.040)	55.87	4.01 (1.150)
55	18.63	3.78 (1.060)	55.58	4.17 (1.180)
56	18.39	3.90 (1.080)	55.39	4.32 (1.210)
57	18.77	3.94 (1.111)	55.87	4.36 (1.240)
58	19.71	3.90 (1.140)	56.96	4.29 (1.290)
59	20.68	3.83 (1.170)	58.18	4.18 (1.330)
60	21.82	3.73 (1.210)	59.48	4.02 (1.370)