Comments Prepared for the

CASAC Ozone Review Panel Teleconference

11/5/2012

Comments on the CASAC Ozone Review Panel's Draft Reports on the First External Review Draft of the Health Risk and Exposure Assessment for Ozone and the Third External Review Draft of the Integrated Science Assessment for Ozone and Related Photochemical Oxidants

William F. McDonnell MD, PhD
William F. McDonnell Consulting
10/31/2012

The EPA with support from CASAC has proposed that a new health risk model for ozone-induced lung function effects be utilized in addition to the previously used model in the upcoming Health Risk and Exposure Assessment (HREA) for Ozone. This new model (MSS model) and its performance have been described in a recent series of published articles (McDonnell et al., 2007, 2010, 2012). My comments below are primarily intended to address specific issues that have been raised about the model in the CASAC review of the HREA and to point out some inconsistencies in the underlying HREA and ISA documents.

Comments on the Draft CASAC Review of the First External Review Draft of the HREA

P. 14, L.23-26 and P. A26, L.16-19. Missing FEV_1 data in the Schelegle et al., (2009) study occur only for measurements made during the filtered air recovery period 4 hours following the end of the 6.6-hr exposure. Approximately ½ of the participants did not complete this final measurement after each of the 5 exposures. Dr. Schelegle indicated to me that the FEV_1 measurement after 4 hours of recovery was originally included in the study only because the investigators were also measuring markers of ozone exposure in exhaled breath condensate at this time. Approximately half way through the study the investigators noted that the biomarkers in exhaled breath condensate were showing no effect of ozone at this time point, and they decided to drop all measurements at this time point for subsequent participants. Thus, the missing data were unrelated to the magnitude of response (personal communication, Edward Schelegle, 2012). The occurrence of missing data does not bias the model-fitting results if the occurrences are attributable to causes that are entirely unrelated to the exposure-response being measured.

A small number of participants in several of the EPA studies of prolonged exposure did not participate in all scheduled exposures. In most cases of missing exposures, individuals simply didn't show up for scheduled exposures, and there is no documentation of the reasons for lack of completion. Although not observed, it is possible that some volunteers selectively dropped out because of discomfort during a

previous ozone exposure. If that kind of drop-out occurred, our estimates of the regression coefficients and the variance components might not be unbiased. My impression is that the large majority of those who did not fully participate did so because of scheduling difficulties. Many college students, the majority of EPA study participants, simply found that they couldn't miss a full day of class or work on multiple occasions, or that they had an upcoming exam, or the opportunity to take an unplanned trip arose, among other reasons common in this population. The results of all exposures, regardless of whether the volunteer participated in all planned exposures, were included in the modeling efforts of the McDonnell et al. (2007, 2010, and 2012) manuscripts. As noted above, if some of the missing exposures are due to "outcome dependent" lack of participation, then bias could exist in the estimates of the model parameters, although to a lesser degree than if all data for those with incomplete participation were excluded. It is unlikely that one will be able to reliably detect instances of "outcome dependent" dropout from examination of the data. It may be worthwhile, however, to document the proportion of volunteers who didn't complete all exposures and to compare the available responses of those who did drop out with those of volunteers who completed all exposures. Similar levels of responsiveness among those with complete and those with incomplete data would suggest less probability of bias.

P. 14, L.26-30 and A-26, L.6-15 and L. 20-24. The issues raised here regarding how results from additional studies with different patterns and levels of concentration and activity will be collated apply only to the Bayesian Markov Chain Monte Carlo modeling approach used in the 2007 HREA for FEV_1 in which observed results from a group of studies with nearly identical exposure patterns and levels of exercise were combined to estimate response at a single time point for the common exposure conditions. In the previous version, this was limited to 6.6-hr exposures with quasi-continuous moderate intensity exercise and a constant ozone concentration for each exposure. Data from new studies includes exposures of longer duration, different activity patterns and levels of exercise, and different patterns of C which complicates direct comparison or combination with one another or with the previous group of studies with common exposure patterns. One of the strengths of the McDonnell Stewart Smith (MSS) model is that exposure and response data from all time points of exposures with a wide range of exposure patterns and levels of C and activity can simultaneously be used to estimate the regression coefficients of the model. We are not limited to using only data from studies with precisely the exposure of interest. This results in much larger datasets and numbers of participants available for parameter estimation. A second strength is that once the coefficients of the model have been estimated, the distribution of predicted population responses can be obtained for any exposure scenario of scientific or regulatory interest by plugging the relevant set of exposure conditions into the model. Assuming that the model fits the original data adequately across the entire range of exposure conditions of interest, the model can generate predictions for exposures of interest which may not have been explicitly studied, but which represent interpolations from large amounts of data from surrounding exposures. It is proposed that for purposes of risk assessment the MSS model be used to generate predicted FEV_1 responses for the individual exposure patterns generated by APEX under various regulatory scenarios of interest. The model can also be directly evaluated in the risk assessment context by forcing APEX to generate a set of

exposures consistent with those of individual chamber studies and comparing the observed with the predicted responses.

P. 15, L. 1-2 and P. A81, L.9-35. As pointed out by Dr. Ultman, it appears that there may be some small differences in how the FEV_1 outcome variable is calculated and how adjustments for filtered air (FA) are made by EPA and in the McDonnell et al. (2012) manuscript. The differences between the methods are likely to result in only small differences in results. For clarity, however, the methods used for the McDonnell Stewart Smith (MSS) model and the rationale for their use are detailed below. We calculated the FEV_1 % decrement for any exposure (including filtered air) at any time point as

$$FEV_1 \text{ \% decrement}(t_i) = 100\% \times [FEV_1(t_i) - FEV_1(\text{pre-exposure})] / FEV_1(\text{pre-exposure})$$

where the pre-exposure value is that value measured on the day of exposure. For every ozone exposure, each individual's FEV_1 % decrement measurement at a given time point was adjusted for FA response at that time point by subtracting the mean of the FA FEV_1 % decrements of all participants in a given study at the same time point. This provides an unbiased estimate of the % decrement in FEV_1 that is due to ozone independent of any effects of the exposure protocol itself (identified as $\%\Delta FEV_1$). Thus,

$$\%\Delta FEV_1(t_i, O_3) = FEV_1 \text{ \% decrement } (t_i, O_3) - \text{mean } FEV_1 \text{ \% decrement } (t_i, FA)$$

An alternate method, apparently used by EPA, involves subtraction of each individual's FA FEV_1 % decrement at a given time point from an ozone FEV_1 % decrement at the same time point (Individual FA adjustment). This also provides an unbiased estimate of $\%\Delta FEV_1$, but has the disadvantage of subtracting a relatively (and occasionally quite) noisy measure (individual FA FEV_1 % decrement) from the ozone FEV_1 % decrement resulting in greater variability (due to noise) in $\%\Delta FEV_1$ for the ozone exposures. If one assumes that, for most healthy individuals, true individual differences in the FA response at any time point are small relative to the variability of the measurement, subtraction of the mean FA FEV_1 % decrement results in less variability in $\%\Delta FEV_1$. To the extent that this assumption is incorrect, use of the mean FA adjustment results in loss of information about individual FA responses and introduces noise into the $\%\Delta FEV_1$. I suspect that these are two competing processes. Use of the mean FA adjustment method also has the advantage of allowing for FA adjustment for studies in which not everyone undergoes a FA exposure either by design or as the result of incomplete participation.

In an informal analysis (unreported data), we compared the results of using both methods of adjusting for FA response (individual FA adjustment and mean FA adjustment) to calculate the $\%\Delta FEV_1$ measured after 5.6 and 6.6 hours of exposure to a constant concentration of 0.08 ppm in two studies from the University of California at Davis (Adams 2003, 2006) and in two studies from EPA (Horstman et al., 1990; McDonnell et al., 1991). Although not independent observations, we calculated the $\%\Delta FEV_1$ for 300 observations. As expected, both methods of correcting for FA produced identical mean $\%\Delta FEV_1$ at each time point of each study. We also tabulated the numbers (% of total) of observations with ozone-induced FEV_1 decrements greater than 10%, 15%, and 20% which are presented in the table below.

FEV_1 Cut-Point	Mean FA Correction	Individual FA Correction
All Studies (n=300 observations)		
>10%	54 (18%)	57 (19%)
>15%	29 (10%)	29 (10%)
>20%	15 (5%)	17 (6%)

The observed numbers with response above a given FEV_1 cut-point were very similar for the two methods of correction. We interpreted the slightly smaller proportions using the Mean FA method as being possibly the result of decreased variability in response due to the method of FA correction. Such differences could also be the result of chance. In any event, in this informal analysis of a small sample it appeared that the two methods of adjusting for FA responses provided very similar results.

Once the FA adjustment is completed for each data point, the filtered air responses are not included in the model fitting for several reasons. The mean of the FA responses for each study, if adjusted for themselves, would be zero at every time point, and the model is written to predict the central tendency of response for a FA exposure to be zero at every time point. The variance of the individual FA exposures, adjusted for the mean FA response, would likely be different from the variance of the $\%\Delta FEV_1$ of very low level ozone exposures which are conducted on a separate day from the mean FA response. Inclusion of the FA responses in model fitting would thus provide no additional information regarding the central tendency of response at zero exposure and would likely provide biased information regarding the variability structure of response at low ozone exposures.

P.A9, L.20 Caution should be used in extending the finding of a threshold in the FEV_1 model (MSS) based upon chamber studies in McDonnell et al., (2012) to other endpoints for at least 3 reasons. First, the MSS model described in 2007 and 2010 and as the "original" model in 2012 is an empirical model developed to best fit a set of chamber data over a wide range of exposure conditions. The behavior of the original model at low levels of exposures is somewhat constrained by observed data at high levels of exposure and doesn't necessarily provide the optimal fit to the data for the lowest level exposures. One purpose of the McDonnell et al. (2012) analysis was to determine whether addition of a threshold to this existing model would improve the fit of the model to the data, particularly for lower exposures, which it did. Before concluding that the existing chamber data imply the existence of a threshold, however, one would need to determine whether other model forms might equally well describe the data without need for a threshold. We did not do that. Second, the mechanistic pathways involved in ozone-induced decrements in FEV_1 are much more complex than indicated in this simple empirical model, and it is generally accepted that other effects such as nonspecific airway hyperreactivity and inflammation are not well correlated with FEV_1 changes suggesting a divergence of mechanistic pathways for different responses. If a threshold in FEV_1 response does exist, we have little current knowledge about where in the mechanistic pathway such a phenomenon resides. It could possibly be in a part of the pathway that is common to multiple endpoints or in a step specific to the FEV_1 response. The mechanistic pathways involved in ozone-induced mortality are even less well understood

than those for lung function changes but are likely to be complex. Existence of a threshold at some point in the FEV_1 pathway does not necessarily imply the existence of a similar threshold in the pathway for mortality. Third, exposures estimated to be below threshold in the chamber data are limited to the earliest time points of the currently available chamber exposures. Although the model would predict that the concentration threshold for a 6.6-hour exposure with moderate exercise is well below 0.04 ppm, this is an extrapolation as no chamber data exist for exposures below 0.04 ppm to directly test such a hypothesis. On the other hand, there are numerous studies of individuals exposed while at rest or undergoing low levels of exercise for short durations in which no changes in FEV_1 were observed. Although these studies generally had low statistical power, they do suggest the possibility of a concentration threshold that is dependent upon activity level and duration of exposure. These findings as well as those of McDonnell et al. (2012) and those of Schelegle et al. (2012) raise interesting questions about the possible existence of a threshold for FEV_1 response and suggest the need for further exploration of the concept for FEV_1 as well as for other ozone effects. At present, however, it is premature to draw any conclusions about the existence of a mechanistic threshold for ozone-induced effects based solely on our current modeling efforts.

P. A82, L.26-29. Some differences in predictions of FEV_1 response between the model used in the 2007 HREA and the MSS model are to be expected given the finer level of detail incorporated for demographic characteristics and exposure conditions in the MSS model and the larger dataset to which the MSS model was fit. However, I strongly agree that it would be worthwhile to better understand any fundamental differences in model performance due to differences in methods of calculating outcome variable, in methods of adjusting for FA, and in the number of times that responses are sampled during a day, among other things. One way to approach this would be to force APEX to generate a set of exposures and demographic characteristics identical to the dataset from which the 2007 HREA model was derived and then use both models to predict responses of this dataset. In the event of unexplained differences between the two models, the observed data from these studies could be consulted if necessary.

Comments on the HREA
(1st External Review Draft)

P. 3-22, L. 29 through P. 3.23, L. 1. I do not believe that this statement is correct with regard to the controlled human studies although it was written in the ISA. See comments below for P. 2-35 Section 2.5.5.4 of ISA. Although a threshold for FEV_1 changes in the controlled human studies has not been identified within the range of concentrations studied for long-duration exposures with moderate exercise, extrapolation of the approximately linear C-R function below the lowest studied level of 0.04 suggests the existence of a nonzero threshold unless the shape of the C-R function changes at low level.

Comments on the ISA
(3rd External Review Draft)

P. 2-35 Section 2.5.5.4 Ozone Concentration Response Relationship
I suggest that an "s" be added to the word "Relationship(s)". There is no reason to expect that all ozone-induced health effects will have mechanistic pathways that substantially overlap, and therefore no reason to expect that they will necessarily have C-R relationships with the same characteristics. Most of this section with the exception of the second paragraph refers to epidemiologic studies and outcomes other than lung function changes. I don't think the first paragraph accurately characterizes the ozone-FEV_1 relationship. The authors might consider moving the second paragraph to the end of the section and there make reference to C-R relationships of lung function changes measured in controlled chamber studies (I would argue that it is the outcome rather than the study venue that is important here).
P. 2-35, L.7. I would suggest that the words "controlled human exposure and" should be removed leaving the sentence to refer only to the epidemiology morbidity and mortality relationships. The following sentences of that paragraph refer to a "generally linear C-R curve with no indication of a threshold" which appears to be a characteristic of the epidemiology mortality studies and studies with other outcomes. Although the data for FEV_1 changes from controlled studies go only as low as 0.04 ppm with no clearly demonstrated threshold after a 6.6 hr exposure with moderate exercise, Figures 6.1A-B show C-R data that are either not linear or suggest the possibility of a threshold at a nonzero concentration. Furthermore, there are plenty of older controlled human exposure studies with exposures conducted for shorter periods of time at rest or low levels of exercise that report exposures to concentrations above 0.04 ppm with no effect on FEV_1 and other physiological outcomes. This at least suggests the possibility of a concentration threshold dependent on activity level and duration of exposure for these outcomes. Individually, each of these older studies typically had little statistical power, and no formal attempt has been made to study the possibility that a true threshold exists for lung function changes. The authors may wish to add a statement referring to such "no observed effect" findings under other exposure conditions to the end of the second paragraph on this same page, and perhaps describe the graphs in Figure 6.1A-B as either being nonlinear or suggesting the possibility of a threshold below the lowest level studied.

P. 6-17 through 6-18. On the second page of the section describing intersubject variability of response, there is much description in the text of proportions of subjects with responses >10% on a study by study basis much of it based on uncorrected (for FA) data. Many of these data (corrected for FA) are summarized in Figures 6 and 7 of McDonnell et al. (2012) both for observed responses and those predicted by the model (note that conclusions are similar to what is now in the text, but the figure provides data at multiple time points in a compact form).

P. 6-21, L.22. When comparing the responses of exercising males and females to ozone in controlled exposure studies, it is necessary to describe whether the activity level resulted in equal levels of V_E or in equal levels of V_E adjusted for some aspect of body

size. In the Hazucha et al. (2003) study cited in the ISA, activity level was chosen such that both men and women had equal levels of V_E/BSA. The same was true in the Adams studies of the 2000's.

REFERENCES

Adams, W.C., 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* 2003; 15:265-281.

Adams, W.C., Comparison of Chamber 6.6-hr Exposures to 0.04-0.08 ppm Ozone via Square-Wave and Triangular Profiles on Pulmonary Responses. *Inhal Toxicol* 2006; 18:127-136.

Hazucha, M.J., Folinsbee, L.J., Bromberg, P.A. Distribution and reproducibility of spirometric response to ozone by gender and age. *J Appl Physiol* 2003; 95:1917-1925.

Horstman, D.H., Folinsbee, L.J., Ives, P.J., Abdul-Salaam, S., McDonnell, W.F. 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 Rev Respir Dis* 1990; 142:1158-1163.

McDonnell, W.F., Kehrl, H.R., Abdul-Salaam, S., Folinsbee, L.J., Devlin, R.B., O'Neil, J.J., Horstman, D.H. Respiratory Response of Humans Exposed to Low Levels of Ozone for 6.6 Hours. *Arch Environ Health* 1991; 46:145-150.

McDonnell, W.F., Stewart, P.W., Smith, M.V. The Temporal Dynamics of Ozone-Induced FEV_1 Changes in Humans: An Exposure-Response Model. *Inhal Toxicol* 2007; 19:483-494.

McDonnell, W.F., Stewart, P.W., Smith, M.V. Prediction of ozone-induced lung function responses in humans. *Inhal Toxicol* 2010; 22:160-168.

McDonnell, W.F., Stewart, P.W., Smith M.V., Kim, C.S., Schelegle, E.S. Prediction of lung function response for populations exposed to a wide range of ozone conditions. *Inhal Toxicol* 2012; 24:619-633.

Schelegle, E.S., Adams, W.C., Walby, W.F., Marion, M.S. Modeling of individual subject ozone exposure response kinetics. *Inhal Toxicol 2012;* 24:401-415.