

Comments Prepared for the  
Clean Air Scientific Advisory Committee  
Ozone Review Panel  
Public Meeting

3/25/2014 – 3/27/2014  
Chapel Hill, NC

Comments on EPA's *Health Risk and Exposure Assessment for Ozone* (Second External Review Draft, February 2014)

William F. McDonnell MD, PhD  
William F. McDonnell Consulting  
3/13/2014

The EPA with support from CASAC has utilized a new health risk model for predicting ozone-induced FEV<sub>1</sub> effects in the current draft of its Health Risk and Exposure Assessment for Ozone (HREA). This new model (MSS model) and its performance have been described in a recent series of published articles (McDonnell et al., 2007, 2010, 2012, 2013). The HREA provides a description and results of a risk assessment for the proportion of individuals experiencing 10%, 15%, and 20% FEV<sub>1</sub> decrements based primarily on the MSS model, and it discusses the influence of various inputs and assumptions on the risk estimates. It also makes comparisons to risk estimates calculated using the health risk model (E-R model) that was used in the 2007 Ozone HREA. The purpose of these comments on the current 2<sup>nd</sup> draft HREA is to identify inconsistencies or inaccuracies in the document, indicate sections that might benefit from clarification, and to discuss characteristics of, and differences between, the two health risk models, some of which may be considered for inclusion in future drafts of the HREA.

### **Comments on Health Risk and Exposure Assessment for Ozone**

#### General Comments

**Comparison of Two Methods.** When conducting risk assessments using two different health risk models, one expects to find differences in estimated risk for the two methods. In order to properly evaluate the risk estimates from a new method relative to an older method, it is useful to understand how various factors influence the risk estimates for each method and how one method may provide biased results relative to the other at least in direction if not in magnitude. Both the older exposure-response (E-R) model and the newer MSS model describe the proportion of a sample of 18-35 year-old volunteers who experienced 10%, 15%, or 20% FEV<sub>1</sub> decrements following 6.6-hr constant-

concentration (square-wave) ozone exposures (range 40 to 120 ppb) while undergoing 5 hours of moderate intensity exercise that required an effective ventilation rate (EVR) of approximately  $20 \text{ l}\cdot\text{min}^{-1}\cdot\text{m}^{-2}$  body surface area. Because the E-R health risk model was based upon the responses of a group of volunteers with a limited range of exposure conditions (response at a single 6.6-hr time point, square-wave exposures ranging from 40-120 ppb, and a single moderate-intensity exercise level), it can be expected to produce accurate risk estimates across a range of 40-120 ppb only when applied to a similar population of individuals with the same distribution of personal characteristics and exposure conditions as it was based upon. In contrast the MSS model accommodates, as input on an individual level, ozone concentration and minute ventilation expressed as functions of time, as well as age and body mass index (BMI), and allows for recovery as well as development of response. This model was fit to data from chamber exposures up to 8 hours in length with FEV<sub>1</sub> measurements recorded hourly, with ozone concentrations varying from 40 ppb to 400 ppb, with exercise levels varying from rest to high intensity exercise, and with varying temporal patterns of ozone concentration including square-wave exposures, triangular concentration pattern exposures, and periods of recovery in clean air following exposure. The model generally fit the observed data across the wide range of exposure conditions. The expectation is that the MSS model should provide accurate predictions of FEV<sub>1</sub> response and therefore accurate risk estimates when combined with estimates of exposure throughout the duration of exposure for a population of individuals exposed in their natural environment under a wide and varying range of exposure conditions. Estimates for populations (such as children and older adults) and exposures beyond the range of data from the original chamber studies reflect extrapolations and greater uncertainty, although existing data from studies of children and older adults suggest that such extrapolations are reasonable.

As is noted in the HREA (e.g. Tables 6-8, 6-9, 6-10), the MSS model consistently provides higher FEV<sub>1</sub> risk estimates than the E-R model when applied to populations across all activity levels. I believe that this is due primarily to underestimation of true risk by the E-R model for a number of reasons, some of which have been identified and discussed in the HREA (e.g. P. 6-30 and Tables 6-9 and 6-10) and which are further discussed below. In addition there are other instances in which the E-R model might be expected to under-predict or over-predict response depending upon the relative contribution of various other factors which are identified and discussed below.

(1) The E-R model is based upon FEV<sub>1</sub> responses measured at the end of 6.6-hr exposures but is applied to 8-hr average ozone concentrations which would contribute to an underestimate of response for a full 8-hr exposure relative to the MSS model, all else being equal.

(2) The majority of the general population experiences activity levels below the range of the chamber studies (EVR =  $13\text{-}23 \text{ l}\cdot\text{min}^{-1}\cdot\text{m}^{-2}$  BSA) upon which the E-R model is based. Because the E-R model cannot be applied to individuals with lower levels of activity, a risk of zero is assumed for these individuals contributing to an underestimate of response for the general population as many

individuals with lower activity levels will experience FEV<sub>1</sub> decrements especially if ozone concentrations are high or for individuals responsive to ozone.

(3) Some individuals may experience periods during the day when ozone concentration and activity level are higher than the 8-hr averages (for example if exercising outdoors for several hours). Such intense, shorter duration exposure could result in maximal FEV<sub>1</sub> decrements predicted to occur during the intense exposure by the MSS model which would not be predicted by the E-R model which is based only upon the 8-hr averages of ozone concentration and EVR. This could contribute to an under-prediction of response for the E-R model relative to the MSS model.

(4) The parameter estimates of the E-R model are based on a sample of study volunteers with a mean age of approximately 23 yrs. Responsiveness to ozone is known to increase with decreasing age over the range of 18-35 yrs. The current RA assumes that the ozone responsiveness of children is equivalent to that of an 18 yr-old which is greater than that of someone 23 yrs of age. Application of the E-R model to children or to any other population with mean age less than 23 yrs will contribute to a smaller estimate of risk than will application of the MSS model which predicts response as a function of age and which is consistent with the assumption of the current RA regarding responsiveness for those less than 18 yrs of age. For populations of adults with mean age greater than approximately 23 yrs, the E-R model will contribute to an over-prediction of risk relative to the MSS model.

(5) The E-R model is based upon exposure of volunteers with a mean EVR of approximately 17 l-min<sup>-1</sup>-m<sup>2</sup> BSA with a range of from 13 to 23. As seen in Figure 6-11 for the Atlanta 2006 base case, ages 18-35 with EVR>13, the median EVR is approximately 14.5 suggesting that application of the E-R model to this population would contribute to an over-prediction of response. However, for children, for whom EVR tends to be much higher (see Figure 6E-19 in Appendix 6E and Pages 5-52 and 5-53 of this 2<sup>nd</sup> draft of the HREA), application of the E-R model to this population would contribute less to an over-prediction of response, or might possibly contribute to an accurate prediction or an under-prediction.

On P. 6-30 and 6-31 and in Tables 6-9 and 6-10, the proportion of individuals, who have an FEV<sub>1</sub> decrement >10% but whose EVR does not exceed 13 L-min<sup>-1</sup>-m<sup>2</sup> at the same time as the FEV<sub>1</sub> response, was estimated. For the Atlanta and Los Angeles cases, the amount of under-estimation of risk by the E-R model explained by this factor (see (2) above) was quantified and accounts for a substantial part of the difference in risk predicted by the two models for this population. In reality, however, all of the above factors make some contribution to under- or over-estimation for any population, and observed differences between the E-R and MSS models reflect the summation of the effects of these factors. The relative importance of the factors that account for differences in risk estimates changes from city to city and year to year and will depend upon the joint distributions of ages, EVRs, concentration patterns, and other factors in

each city for each period of time. Although the curves in Figure 6-10 suggest similar predicted responses for both models when restricted to  $EVR > 13$  for the Atlanta base case, 2006, ages 18-35, I suggest that the curves representing children with  $EVR > 13$  would differ considerably for the two models. I suspect that the MSS curve would be considerably higher than the E-R curve due to factors (4) and (5) above.

### Specific Comments

**1. Executive Summary, Figure and text on P. ES-5.** Do the bar graphs represent “average percent increases ...” or “Percent of all school-age...”? Captions do not agree (Same for figure on P. ES-7). The text indicates that the focus is on children, yet in the same paragraph results for “any study group” are intertwined with the results for children. This is a confusing presentation. The text for the graph on page ES-7 is much clearer.

**2. P. 6-8, L. 20-27.** This description of the data used in developing the MSS model is not accurate or is at least confusing. The original MSS model (without the threshold) that was described in the 2007 and 2010 McDonnell et al. papers was developed from a set of 15 controlled exposure studies that are well described and referenced in the 2007 paper. Only 3 of those 15 studies are included (I wouldn't say described) in the table in section 6.2.5. Most of the studies listed in section 6.2.5 were not used in the original development of the model (All were, however, used to estimate parameters in the threshold model in the 2012 MS). The other 12 of the original 15 studies included some 6.6 hr studies at lower levels of exercise as well as shorter duration studies at various exercise levels (only the short duration studies are described in the 1997 paper). I would suggest only referencing the 2007 paper for clarity. In the 2012 paper we refit the 2007/2010 model which now included a threshold to data from the original 15 studies plus data from 8 additional studies (the Adams, Kim, Schelegle studies). These 8 additional studies are well described in the 2012 paper.

**3. P. 6-8, L.29.** Although the terms “accumulated dose” or “net accumulated dose” require some assumptions about the meaning of the variable X in the model, this seems to me to be a more intuitive and descriptive term than “cumulative dose rate” when describing the threshold. The threshold is described in terms of the variable X in compartment 1 of figure 6.1 which can loosely be interpreted as accumulated dose taking into account reductions in X due to metabolism, diffusion, etc. according to first order reaction kinetics. See also **P.6-9, L. 9 and L 12.**

**4. P. 6-10, L 18-20.** In one place you use the expression  $L \text{ min}^{-1}$  for  $V_e$  and in another you express  $V_e/BSA$  as  $L/\text{min}\cdot\text{m}^2$ . The form of these should be consistent for clarity. Also check figure legends for consistency with whatever form you choose.

**5. P. 6-11, L. 1.** This sentence might be more clearly expressed as “The values of the  $\beta_s$  and the variances of  $\{U_i\}$  and  $\{\epsilon_{ijk}\}$  were estimated from fits of the model to the data (see ...)”

**6. P. 6-11, L.3-4.** Were the values of both  $U$  and  $\epsilon$  constrained to be within 2 SDs of the means? Either here or later describe the truncation method in more detail. For example, are values selected from the normal distribution and those above 2 SD then discarded and reselected or those above 2 SD assigned a value of 2 SD?

**7. P. 6-12, L. 16.** This sentence might be clearer as "...we reparameterize the age term in the numerator of equation 6-3 by [...]"

**8. P. 6-12, L.27, 29, 30.** The term "subject" should probably be replaced with "volunteer".

**9. P. 6-15, Figure 6-4.** Is it possible that ozone exposure started at time 1 hr rather than time 0? Otherwise, can I assume that the predicted median  $FEV_1$  response is so low for the first hour of exposure without exercise in the non-threshold model that it is indistinguishable from zero (<0.05%)? The way that the figure is drawn, it appears that the non-threshold model also has a threshold which is not exceeded during the first hour of exposure.

**10. P. 6-15, Figure 6-5.** What is the explanation for the predicted response rate (for  $FEV_1 > 10\%$ ) of 0.8% for ozone exposure below threshold? Were the possible values of epsilon not capped at twice the SD as indicated on page 6-11? Although this is a small proportion, if APEX runs predict nonzero probabilities of response at zero ozone concentration (which would not be inconsistent with the observed chamber data), this should be noted when discussing the proportions of the population and the raw numbers experiencing effects in the risk assessment. Such an effect could be stronger than any true effect of ozone at very low concentrations.

**11. P. 6-16, L.11-13.** This sentence should be changed to read "The controlled human exposure study data were corrected on an individual basis for study effects in clean filtered air to remove any systemic bias that might be present in the data attributable to the effects of the study itself (e.g. exercise, diurnal effect, etc.) (ISA, Section 6.2.1.1)." Note that effects observed in the clean air exposure may have causes other than those due to exercise alone.

**12. P. 6-16 and 6-17, Table 6-3, Figure 6-6 and text.** In Table 6-3, numbers of participants with  $FEV_1$  changes at end-exposure are given for each study as are the numbers with  $FEV_1$  changes at any time during exposure (in parentheses). It is my impression that the numbers with  $FEV_1$  changes at end-exposure are the ones modeled. Please clarify this in the text.

**13. P. 6-26, Figure 6-9.** Although not of critical importance, it would be interesting to know the circumstances under which the 1% of all  $FEV_1$  decrements occurred when ozone was between 20 and 30 ppb and the 4% occurred when ozone was between 30 and 40 ppb. Is the presence of decrements at such low concentrations the result of selecting a large epsilon from its distribution which would result in predicted responses at near zero

exposure (see comment 10 above) or are there other exposure factors (such as high EVR for prolonged periods, or a very large number of exposures to these low levels) at play?

**14. P. 6-26, L. 17-20 and Tables 6-4 and 6-5.** I find it interesting that the proportions of children with FEV<sub>1</sub> changes decrease substantially when going from one event per year to 6 events per year. (A similar finding was presented in the Policy Assessment Figures 3-11 through 3-14 for two events per year and perhaps should be referenced here.) My initial impression was that a single FEV<sub>1</sub> event might represent occasional selection of a large value for epsilon. However, Figures 3-7 through 3-10 in the Policy Assessment indicate that the numbers of exposures of concern also drop dramatically when going from 1 to 2 events per year. This suggests that the drop in the number of individuals experiencing multiple FEV<sub>1</sub> events compared to the number experiencing a single event is driven by a single large, possibly atypical, exposure for each individual rather than to an attribute of the model (namely selection of a value of epsilon from the distribution of possible values). It seems to me that presentation of data for 1 or more events, 2 or more events, and 6 or more events gives a more complete picture of how meeting the current standard and alternate standards affects risk.

**15. P.6-33, L. 2-4.** The text indicates a higher MSS response for the low and high ranges of concentration. Figure 6-10 indicates a higher E-R response for the low and high ranges. These should be consistent.

**16. P. 6-38, L. 11-12.** I would suggest rewording the sentence to read: “(1) FEV<sub>1</sub> responses increase with increasing O<sub>3</sub> concentration, ventilation rate, and duration of exposure if the other two variables are held constant.” That is, response doesn’t necessarily increase with duration if ozone concentration is also falling with duration.

**17. P. 6-38, L. 24 through 6-39, L. 7-10 and Figure 6-12.** We have made observations that suggest that uncertainty may be somewhat less than projected by the analysis presented, at least when the model is applied to the dataset from which it was derived. The parameters of this model are correlated, and the model is quite flexible. That is, we have on occasion fixed a coefficient at a value different from its best-fit value and refit the model. Some of the new estimates of the coefficients change in response to the new value of the fixed coefficient, and the predicted population responses for the two models are generally very similar. Predicted values for some individuals do change, but the overall predicted population responses are generally unaffected. One example of this can be seen in Table 6-1 and Figure 6-5. The threshold model in Table 6-1 (labeled 2012T) allowed  $\beta_9$  to be estimated with the resultant value being 59. The non-threshold model in Table 6-1 (labeled 2012) reduced the value of the threshold ( $\beta_9$ ) such that it is fixed at zero and the model was refit. To compensate for this, the  $\beta_4$  term in the refit model changed substantially, and the  $\beta_1$  term changed somewhat. Figure 6-5 shows that the probability of response for the study sample was essentially unchanged by this rather large change in parameter values. Again, this is an example of a situation in which the model is applied to the sample from which it was derived, and individual predictions may change somewhat when one parameter is changed. If models with two different sets of coefficient values were applied to cities and time periods with characteristics of

individuals and exposures that were different than the experimental data set, it is possible that the probabilities of response would differ for the two sets of values. However, one should keep in mind that compensatory changes can occur in other model coefficients that can reduce the sensitivity of the population responses to uncertainty in coefficient values.

**18. P. 6-39, L. 11-19.** I agree that there is substantial uncertainty regarding the proper value of the age term and extrapolation of the model down to age 5 yr. As noted above, because the model is flexible, fixing the value of the age coefficient at a different value from the best fit value has little effect on the population predictions for the sample to which it is fit although it changes the predictions for some individuals depending on their age. If one applies the model with two different sets of coefficients to a population with a different age distribution (e.g. children), it is unlikely the two models will give the same population predictions. When the model was originally fit to the EPA data set (n=540) (McDonnell et al, 2007), the age coefficient ( $\beta_2$ ) was a highly statistically significant -0.41. When the U.C. Davis data (n=142) were added and the model was refit, it became -0.19 which was not statistically significant. When the U.C. Davis and EPA data were allowed to have separate age coefficients with other coefficients in common (McDonnell et al 2012), we found a value for the U.C. Davis age coefficient of +0.19 and for the EPA data, -0.45. We observed that in the U.C. Davis data, age was correlated with body mass index (BMI), while there was no correlation in the EPA data. Adding BMI to the combined data set resulted in a still nonsignificant value of -0.26 for the combined data. In our McDonnell et al (2012) we noted that Bennett et al (2007) had analyzed the effect of age on ozone response in 197 individuals, ages 18-35. They found a response that was consistent with a value of approximately -0.25 for an age coefficient. To further complicate things, we recently fit the combined EPA and U.C. Davis data (McDonnell et al. 2013) using a different structure for within-subject variability, and found an age coefficient of -0.43 which was statistically significant. Based on the EPA and Bennett data and the numbers of participants in those studies, and the new analysis with a different variability structure, I feel confident that the correct age coefficient is in the -.20 to -.45 range. We have not fit the U.C. Davis data alone with the new variability structure model. At this point it is difficult to explain the positive age coefficient for the U.C. Davis data.

**19. P. 6-41, L. 1-14.** Within-individual variation in response ( $\epsilon$ ) captures variability in FEV<sub>1</sub> measures both between exposure days and within a single day. It is my opinion that this term reflects real day-to-day variability in responsiveness to ozone as well as pure error. If the term reflected only error,  $\epsilon$  could be set to zero, and the model would provide predictions of central tendency of response for each individual for each exposure which would deviate from the observed data (which contains error), but which might be a more precise measure of the effect of ozone. Because the term also contains information about true variability in response (which probably occurs between days), it cannot be ignored when generating predictions of response. In the 2012 MSS model, the variance of  $\epsilon$  was assumed to be normally distributed and constant for all levels of exposure and was estimated from the model fit. To generate predictions of FEV<sub>1</sub> response for simulated exposures, a value of  $\epsilon$  is selected from the distribution of possible values once

a day for each individual exposed and added to the individual predicted central tendency of response. Because the estimated variance of  $\epsilon$  is rather large (17.1, SD = 4.1) relative to a 10% FEV<sub>1</sub> decrement, the probability that an individual will experience a 10% FEV<sub>1</sub> decrement is quite sensitive to the value of  $\epsilon$  selected. EPA has decided, based on conventions previously used in APEX, to truncate the possible values of  $\epsilon$  at +/- 2 standard deviations from its mean of zero (+/- 8.27) to reduce the sensitivity of results to extreme values of  $\epsilon$ . This should reduce the proportions of individuals with low level exposures (and possibly higher level exposures) experiencing FEV<sub>1</sub> effects, consistent with the results of Table 6F-4 of Appendix 6F.

Additional information suggests that this truncation may not be arbitrary, at least for low level exposures. The original 2012 MSS model assumed the variance of  $\epsilon$  to be constant for all levels of predicted response. Because  $\epsilon$  contains information about day-to-day changes in ozone responsiveness, and because differences in ozone responsiveness manifest themselves to a greater extent when the central tendency of response is predicted to be large (e.g. high exposures or highly responsive individuals), it seems that the variance of  $\epsilon$  should be lower for smaller predicted responses and higher for larger predicted responses (See McDonnell et al 2013 for further discussion). We therefore refit the MSS model using a new variance structure in which the variance of  $\epsilon$  was assumed to be related to the magnitude of predicted central tendency of response (McDonnell et al., 2013). This model fit the data better than the original model that assumed constant variance (although predicted population responses in the original data set were very similar for both models probably because of the shifts observed in other coefficients). The estimated within-individual variance of response for exposures below threshold was 9.1 in the new model rather than the constant 17.1 in the previous model. This value of 9.1 was consistent with the variance of observed responses for below threshold exposures. The level at which  $\epsilon$  is truncated in the HREA (+/-8.27) is close to 3 standard deviations from the mean for low-level exposures as estimated in the new model. For exposures and individuals with small predicted responses, the smaller selected values of  $\epsilon$  for low level exposures result in more precise estimates of FEV<sub>1</sub> decrement for each individual and therefore more accurate estimates of the proportion experiencing a 10% FEV<sub>1</sub> change. In the new model the predicted variance of  $\epsilon$  increases with increasing predicted response. We have not explored the effect of increasing variance in  $\epsilon$  on predicted response for exposures with higher predicted responses nor is it clear how truncation would affect predicted responses at higher levels of exposure.

An alternate approach to use of the constant variance model with truncation of the possible values of  $\epsilon$  would involve selection of values of  $\epsilon$  from the distributions of  $\epsilon$  based upon the proportional variance model, either with or without truncation. This approach could also form the basis for further sensitivity analysis of the influence of  $\epsilon$ . My hunch is that predicted proportions of individuals experiencing FEV<sub>1</sub> effects with low exposures will be somewhat lower. It is not clear to me how an increasing variance will affect predicted proportions with FEV<sub>1</sub> effects as the predicted central tendency of FEV<sub>1</sub> response approaches 10%.

**20. P. 6-46, L. 11.** This sentence could be expanded to the following: "...wide range of exposure times and levels of exercise as well as for exposures with time varying exposures and levels of exercise (Section..."

**21. P. 6-46, L. 31 through P. 6-47, L. 8 and Figures 6-48 and 6-49.** It would be useful to see these two figures for percent of children with 2 or more or 6 or more episodes of FEV<sub>1</sub> response for the various standard levels.

**22. P. 6D-3, Figure 6D-1 and Figure 6D-2.** In the figure caption, "Subject id" should probably be something like "Subject count" as the numbers in the figure are consecutive whereas the true subject numbers in Table 6D-1 occasionally skip a number.

## REFERENCES

Bennett, W.D., Hazucha, M.J., Folinsbee, L.J., Bromberg, P.A., Kissling, G.E., London, S.J. Acute pulmonary function response to ozone in young adults as a function of body mass index. *Inhal Toxicol* 2007; 19:1147-1154.

McDonnell, W.F., Stewart, P.W., Smith, M.V. The Temporal Dynamics of Ozone-Induced FEV<sub>1</sub> 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.

McDonnell, W.F., Stewart, P.W., Smith, M.V. Ozone exposure-response model for lung function changes: an alternate variability structure. *Inhal Toxicol* 2013; 25:348-353.