Comments on First Draft of the NO_2 "Risk and Exposure Assessment"
– Summary of Key Issues –

Prepared on behalf of UARG
for CASAC Meeting of May 1-2, 2008

Anne E. Smith, Ph.D.
CRA International
April 30, 2008

Following is a brief summary of the key issues that we hope CASAC and EPA will consider for improving the usefulness of the NO_2 "Risk and Exposure Assessment" (REA). A longer exposition of these points will be submitted to EPA by the comments deadline of May 30, 2008.

Synopsis: The REA does not provide the coherent structure for supporting policy decisions that is the purpose of a risk assessment, for the following reasons:

- There is no connection between the particular benchmark levels against which exceedances are calculated and what such exposures mean for "risk."
- The 200 ppb benchmark in particular has no apparent justification in the science.
- The REA focuses only on worst case locations and fails to provide a representative characterization of U.S. risks at as-is NO_2 levels.
- The roll up of as-is concentrations to simulate of hypothetical exposure scenarios stretches beyond the bounds of realism and is unsupportable statistically.
- EPA does not provide evidence to demonstrate that the Air Quality Characterization and the Exposure Modeling portions of the REA are relying on consistent estimates of as-is NO_2 concentrations.

The net effect is that the results presented in the REA lack "utility."

1. Risk-relevant interpretation of exposure benchmarks is needed.
EPA acknowledges that it cannot construct a meaningful quantitative exposure-response or dose-response curve for NO_2 risk assessment. In its place, it considers exceedances of several "benchmark" exposure levels for NO_2. (Specifically, the benchmarks used in the draft REA are 1-hour concentrations of 200 ppb, 250 ppb and 300 ppb.) Use of benchmarks can be a useful alternative approach *as long as it is possible to link exposures of those benchmarks to observations of specific health effects of concern*.

The problem with the draft NO_2 REA is that it provides no clear risk-related interpretation of exposures that exceed each of the three benchmark levels. Further, the scientific evidence provided in the REA and the "Integrated Science Assessment" (ISA)

appears not to provide justification for using a benchmark of 200 ppb as a point where heightened health effects have been observed to occur.

The REA states that the benchmarks are intended to reflect exposures at which asthmatics are at heightened risk of airways responsiveness.[1] EPA then states that the benchmarks of 200 ppb, 250 ppb and 300 ppb reflect the "lower- middle- and upper-end of the range identified in the draft ISA as the lowest levels at which controlled human exposure studies have provided sufficient evidence for the occurrence of NO_2-related airway responsiveness."[2] Table 1 on p. 19 of the REA is provided to support this statement, by identifying the "key controlled human exposure studies identified in the draft ISA that evaluated airways responsiveness."[3] (A copy of Table 1 is provided at the back of these comments.)

Interestingly, the REA's Table 1 shows that the lowest NO_2 exposure level at which a statistically significant airways responsiveness effect has been detected in asthmatics is 250 ppb. The two studies listed that considered lower levels (*i.e.*, Jenkins, 1999 and Roger 1990) do not find a statistically significant effect, even though those studies actually used longer than 30-minute exposure times. Further, Table 1 shows that statistically significant airways responsiveness effects are not found with increasing likelihood as studies consider exposure levels higher than 250, including up to 600 ppb. Thus, the data provided in the REA only support 250 ppb as a possible "lower-end" benchmark exposure for which studies have provided sufficient evidence for the occurrence of NO_2-related airway responsiveness. In fact, even benchmarks above 300 ppb would fit the characterization that "only a subset of asthmatic individuals exposed at and above a given benchmark level would actually be expected to experience any such potential adverse health effects."[4]

The ISA does not appear to contain the particular listing of studies shown in Table 1, which also raises questions about whether the REA has been sufficiently integrated with the ISA. The summary chapter of the draft ISA (chapter 5) does mention the range of 200 to 300 ppb for observed increases in non-specific airways responsiveness,[5] but the detailed discussion in Chapter 3 of the ISA does not support this range any better than Table 1 of the REA.[6] In fact, the detailed discussion of airways responsiveness studies in Chapter 3 of the ISA only mentions two studies in which exposures were below 260 ppb. One is Roger *et al.* (1990) which did not find an effect at 150 ppb (this appears in Table 1). The other is Tunnicliffe *et al*. (1994), which found an effect at 400 ppb, but not at 100 ppb. Table 1 of the REA only mentions the exposure of 400 ppb at which an effect *was* observed in that study. Thus, Chapter 3 of the ISA appears to only further

[1] "The focus for purposes of characterizing risks associated with ambient NO_2 is airways responsiveness." (REA, p. 15, lines 26-27).
[2] REA, p. 18, lines 6-9.
[3] REA, p. 18, lines 1-2.
[4] REA, p. 18, lines 18-20.
[5] ISA, p. 5-13.
[6] ISA, pp. 3-15 to 3-20.

support the evidence suggest in Table 1 of the REA that there have been no effects found at exposures below 250 ppb, even for airways responsiveness in asthmatics.

In brief, neither the REA nor the ISA offers evidence that benchmarks below 250 ppb have any relationship to health risk, even if risk is defined as a statistically significant degree of airways responsiveness observed in asthmatics in some but not all controlled clinical studies. Risk characterization is supposed to link a health hazard identified scientifically to exposures to estimate the frequency and severity of such effects in populations at large. The REA has not linked the benchmarks it is using to the underlying science. REA estimates of possible population exposures to a benchmark level of 200 ppb would appear to have no coherent risk interpretation founded in the science described in the ISA, and should be removed from the REA. According to Table 1, even exposures above 250 ppb and up to 500 ppb produce inconsistent findings of effect among sensitive individuals and substantial qualification needs to be added to the REA for exposure estimates associated with exceedances of the higher benchmark levels that EPA is using.

2. Selected locations need to be more representative.

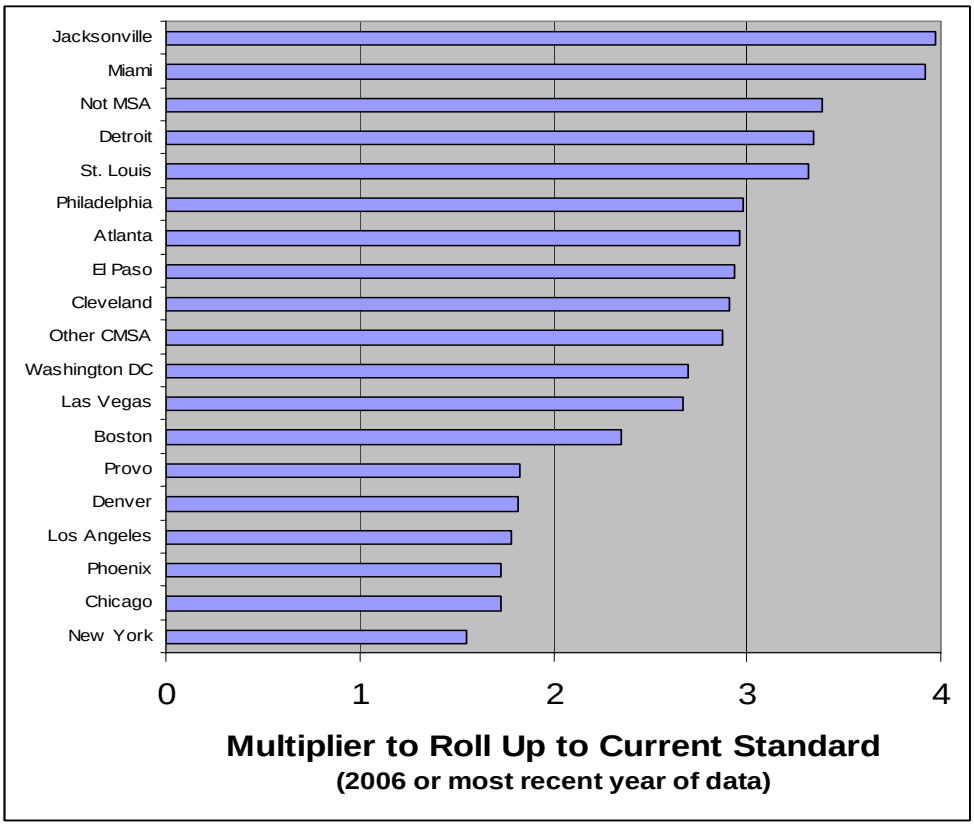
The purpose of a risk assessment is to characterize risks at exposures currently being experienced. A comprehensive understanding of current risks to US residents therefore requires consideration of a range of exposure conditions. Preferably the risk assessment would also inform its readers on the portions of the population at various points along the range of exposure conditions. The REA is not characterizing exposure information in this representative manner.

EPA has applied selection criteria to identify cities ("locations") to study in detail for both the Exposure Modeling and the Air Quality Characterization portions of the REA. Briefly, the selection criteria for the Air Quality Characterization portion require that the city in question (i) have had a monitor whose annual average was among the worst 10% of all monitors in the US for some year, or (ii) had at least one monitored hour that exceeded 200 ppb. The criteria are even more selective of extreme-case conditions in the Exposure Modeling portion of the REA, in that both (i) and (ii) must be met. These selection criteria are inappropriate because they are tailored to ensure that only the worst case exposure conditions are assessed. The REA thus does not provide a characterization of NO_2 exposures that are representative of what is being experienced across the US as a whole.

By providing exposure characterization for multiple cities (*e.g.*, 18 in the Air Quality portion), yet all of them having among the worst 10% of conditions in the US, the REA is likely to convey the biased impression that such high exposures are representative of those across most of the US. The REA should consider a balanced mix of locations that, as a group, reflect the entire range of conditions in urban areas of the US. It should also offer some information on the relative portions of the US population living in areas that are typical of each of the locations on the spectrum.

3. Simulations of exposures at the current standard ("roll ups") are not credible. All of the as-is concentrations in the record of monitoring data are far below the current NO_2 NAAQS of 53 ppm annual average. The REA "rolls up" these as-is monitored air concentrations in an effort to simulate risks if the current standard were just being met. This step requires such excessive extrapolation for NO_2 that it cannot be considered valid, even as a hypothetical exercise. This step of the REA should be abandoned.

The inappropriateness of the roll up procedure starts with the fact that it requires a very extreme degree of extrapolation beyond the data record. The figure below shows the multiples by which EPA is increasing "as is" to simulate air quality at the current NAAQS of 53 ppb.[7] *Except for a few cities, EPA is multiplying the monitored concentrations in every hour of the year by a factor of 2 to 4.* The fact that EPA is doubling to quadrupling air concentrations highlights the absurdity of this hypothetical worsening of US air quality that the current standard might allow.



The inappropriateness of the roll up procedure is further confirmed by information provided in Appendix D of the Technical Support Document (TSD). Appendix D of the TSD reports the highly unsatisfactory results of attempts estimate a statistical relationship

[7] The multipliers in the figure are the "*F*" values found in Table 11 of the Technical Support Document, p. 26. The figure uses the multiplier that is applied to the most recent year of data available for each location.

between the annual means of the monitored NO_2 hourly concentrations and the numbers of hours in the distribution that exceed various peak levels. The statisticians had so little success in predicting exceedances of the lowest peak level that they considered (*i.e.*, 150 ppb) that they did not even attempt to predict exceedances for peaks that occur even less frequently. In other words, they found no reliable relationship between the level of mean concentrations and the frequency of peak concentrations.

Given this statistical failure, EPA fell back to what it calls an "empirical" approach to simulate exposures at alternative annual average NAAQS, which is simply to roll up each individual observation in the as-is data by the same proportion that the mean needs to be increased. This approach may avoid using a formula that has demonstrably invalid statistical properties, but it cannot circumvent the fact that the same problems of extrapolation exist even if the roll up is performed mechanically on each data point in the record. The "empirical" roll up procedure is invalid for the same reasons that the statistical effort to find a predictive equation failed: there is no predictable correlation between the observed mean values and the observed frequency of peak values for NO_2 – yet the roll up procedure assumes that it is completely predictable. Even if mean NO_2 *were* to increase, there is no reliable way to determine how the upper tail of the air concentrations would change. Yet the roll up procedure simply assumes that every monitored value along the distribution would increase by exactly the same percentage as the change in the mean. In these circumstances, a roll up procedure cannot possibly be justified for extrapolations involving multiplying by a factor of 2 to 4.

In my comments on the proposed REA Plan (October 23, 2007), I suggested that if EPA were to roll up concentrations to try to simulate risks at the current NAAQS, it should emphasize how unlikely such an exposure scenario would be under the current Clean Air Act (especially its provisions for Prevention of Significant Deterioration). In light of the additional information presented in the First Draft of the REA regarding (1) the very large roll up factors that are necessary and (2) other evidence provided in the TSD about the difficulties of making any extrapolations with these NO_2 data, the concept of simulating the current standard should now be dropped altogether. Results regarding exceedances of benchmarks under a hypothetical "attainment" of the current standard are unreliable to the point of having no utility for decision makers. This is a problem for both the Air Quality Characterization and for the Exposure Modeling elements of the REA.

4. Exposure Modeling and Air Quality Characterization need to be mutually consistent.

The air concentration estimates used in the Exposure Modeling appear inconsistent with the monitored air concentration information used in the Air Quality portion of the REA. The chance that these two parts of the REA are fundamentally inconsistent with each other is a serious concern that should be investigated closely before relying on results from either or both parts. Much more information should be provided to demonstrate that the exposure modeling results are actually building on and complementing the air quality assessment results.

Table 26 on p. 91 of the REA shows that large, *ad hoc* "adjustments" are being made to the modeled air concentrations in order to produce average air concentrations that are similar to those that have been monitored. (A copy of Table 26 is pasted below.) This adjustment procedure results in very large and crude changes. Notice, for example, that the adjustment adds about 30% more to the raw model-projected concentrations in Table 26. In fact, this adjustment procedure is so gross that it actually reverses which sites in Philadelphia have higher or lower concentrations as compared to their order observed in the monitoring data. For example, the middle of the three monitors listed in Table 26 becomes the worst-case monitor in two of the three years modeled. Similarly, in the 2003 data, the model overshoots the level of the third monitor by 1 ppb, but the adjustment procedure then actually *adds* 6 more ppb to that monitor's value, causing the overshoot to be increased from 1 ppb to 7 ppb.

<u>Copy of Table 26 from p. 91 of the REA</u>

Table 26. Comparison of ambient monitoring and AERMOD predicted NO2 concentrations.

Year and Monitor ID	Annual Average NO_2 concentration (ppb)			
	Monitor	AERMOD Initial	Difference[1]	AERMOD Final[2]
2001				
4210100043	26	7	18	19
4210100292	28	22	6	33
4210100471	30	20	10	32
mean			11	
2002				
4210100043	24	7	17	18
4210100292	28	21	7	32
4210100471	29	19	10	31
mean			11	
2003				
4210100043	24	7	17	13
4210100292	25	22	3	28
4210100471*	25	26	-1	32
mean			6	

[1] the difference represents concentrations attributed to sources not modeled by AERMOD and model error.
[2] the mean difference between measured and modeled was added uniformly at each receptor hourly concentration to generate the AERMOD final concentrations.
* monitor did not meet completeness criteria used in the air quality characterization.

Adjustments like this could dramatically alter estimated exposure patterns, since some sites will be more relevant to time-weighted exposures than others. Huge biases could be created in this adjustment step, and could go undetected, without more careful exploration and comparison of monitored and modeled air concentrations.

At a minimum, EPA should provide a clear comparison of the air concentration data in its Exposure Modeling and that used in the Air Quality Characterization step. This could be done by producing tables comparable to Tables 4, 5 and 6 of the REA but using the air concentration data that is embedded in the Exposure Modeling. This would require using only the Exposure Modeling concentrations for the sites that reflect the location of the monitors that were used in producing Tables 4 through 6. If these prove to be similar, then one can have more confidence that the two parts of the REA are producing complementary results, and that there are no hidden biases due to inconsistencies between the two sets of air concentration data.

Table 1. Summary of Key Controlled Human Exposure Studies of Airways Responsiveness

1

Study	NO_2 Exposure Level (ppm)	Exposure Duration	Study Population	Allergen versus non-specific	Metric Used	Number of Subjects	Exercise	Statistically Significant	Statistically Non Significant
Tunnicliffe, 1994	0.4	1-hour	Mild asthmatics	Allergen	Lung function	8	No	**X**	
Devalia, 1994	0.4	6-hours	Mild asthmatics	Allergen	Lung function	8	No		**X**
Strand, 1997	0.26	30-minutes	Mild asthmatics	Allergen	Lung function	18	No	**X**	
Strand, 1998	0.26	30-minutes (4x per day)	Mild to Moderate asthmatics	Allergen	Lung function	16	No	**X**	
Barck, 2005	0.26	15-minutes (3x over 2 days)	Mild asthmatics	Allergen	Lung function	18	No		**X**
Barck, 2005	0.26	15-minutes (3x over 2 days)	Mild asthmatics	Allergen	Inflammatory Markers (sputum, blood)	18	No	**X**	
Barck, 2002	0.26	30-minutes	Mild asthmatics	Allergen	Inflammatory Markers (BAL)	13	No	**X**	
Bylin, 1985	0.3	20-minutes	Mild asthmatics	Non-specific	Lung function	8	No	**X**	
Mohsenin, 1987	0.5	1-hour	Asthmatics	Non-specific	Lung function	8	No	**X**	
Strand, 1996	0.26	30-minutes	Mild asthmatics	Non-specific	Lung function	19	No	**X**	
Jörres, 1990	0.25	30-minutes	Mild asthmatics	Non-specific	Lung function	14	No	**X**	
Rubenstein, 1990	0.3	30-minutes	Asthmatics	Non-specific	Lung function	9	Yes		**X**
Jörres, 1991	0.25	30-minutes	Mild asthmatics	Non-specific	Lung function	11	Yes		**X**
Witten, 2005	0.4	3-hours	Mild asthmatics	Allergen	Inflammatory Markers (sputum)	15	Yes		**X**
Jörres, 1991	0.25	30-minutes	Mild asthmatics	Non-specific	Lung Function	11	Yes		**X**
Jenkins, 1999	0.4	3-hours	Mild asthmatics	Allergen	Lung function	11	Yes	**X**	
Jenkins, 1999	0.2	6-hours	Mild asthmatics	Allergen	Lung function	11	Yes		**X**
Witten, 2005	0.4	3-hours	Mild asthmatics	Allergen	Lung function	15	Yes		**X**
Roger, 1990	0.15-0.6	75-minutes	Mild asthmatics	Non-specific	Lung function	21	Yes		**X**