



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

June 25, 2008

EPA-CASAC-08-014

The Honorable Stephen L. Johnson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Clean Air Scientific Advisory Committee's (CASAC) Peer Review of
EPA's Risk and Exposure Assessment to Support the Review of the NO₂
Primary National Ambient Air Quality Standard: First Draft

Dear Administrator Johnson:

The Clean Air Scientific Advisory Committee (CASAC), augmented by subject-matter-experts to form the CASAC Oxides of Nitrogen Primary National Ambient Air Quality Standards (NAAQS) Review Panel (hereafter referred to as the panel, roster provided in Enclosure A) held a public meeting on May 1-2, 2008 to review EPA's *Risk and Exposure Assessment to Support the Review of the NO₂ Primary National Ambient Air Quality Standard: First Draft*. The Chartered CASAC held a public teleconference on June 11, 2008 to review and approve the report.

Overall, CASAC found that the first draft of the *Risk and Exposure Assessment* (REA) represents a good start in the development of the document for use in rule-making, but requires major revisions. CASAC found the health evidence presented on acute indicators of risk supports consideration of a short-term (less than or equal to a 24-hour) NO₂ standard. A broader discussion of "at-risk" subpopulations is needed. Other concerns are that the overall scientific evidence for health risks from longer-term exposures is more compelling than its characterization in the current draft. Also, while acknowledging the methodological challenges, the exclusion of epidemiologic exposure/response relationships from the risk assessment is viewed by many as a serious shortcoming of the current draft.

The CASAC had many suggestions for strengthening the document and those suggestions are listed below in the form of answers to the Agency's charge questions. CASAC panel members have also provided individual comments on the document.

Air Quality Information and Analysis

Agency charge questions:

1. **To what extent are the air quality characterizations and analyses technically sound, clearly communicated, appropriately characterized, and relevant to the review of the primary NO₂ NAAQS?**
2. **To what extent are the properties of ambient NO₂ appropriately characterized including ambient levels, spatial and temporal patterns, and relationships between ambient NO₂ and human exposure?**
3. **We have evaluated air quality in a number of individual locations throughout the United States. What are the views of the panel regarding the appropriateness of these locations and on the approach used to select them?**
4. **In order to simulate just meeting the current standard, we have rolled up NO₂ air quality levels. To what extent is the approach taken technically sound, clearly communicated, and appropriately characterized? Do Panel members have comments on the relevance of this simulation for reviewing the primary NO₂ NAAQS?**
5. **What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability?**

Response: CASAC thought that this was a good start toward developing a risk/exposure assessment document, but in a number of ways it was substantially incomplete. Staff appears to have restricted their analytic approaches too early in the process. The CASAC recommends the inclusion of a description of the overall Risk & Exposure Assessment (REA) approach at the beginning of the chapter. Use of flowcharts depicting how the different models are used and provide inputs to other models with supporting text would be useful. The equations for roll-back of the health effect benchmarks should be provided and their limitations discussed (e.g., assumption of linearity).

The approach for calculating the on-road concentrations is based on an empirical relationship with parameters derived from published monitoring studies conducted at various distances from roadways. It would add scientific credibility to this study to conduct an evaluation of this approach using an independent data set. For example, the maximum nitrogen dioxide (NO₂) concentration may not necessarily occur on the roadway because NO will become oxidized to NO₂ as the roadway atmosphere becomes dispersed and mixes with the background ozone.

For estimation of exceedances, the largest values (upper tail) of the NO₂ concentration distribution are the most important. Evaluation of the use of the air quality data must be done to determine whether the current approach is reasonable. The extreme of the NO₂ concentration distributions may occur in configurations such as street canyons that are not treated in the current analysis. If it is not possible to address such extreme situations in the current framework, this limitation should be explicitly stated and its implications on the uncertainties of the results should be discussed [see individual comments of Dr. Elizabeth A (Lianne) Sheppard].

It is important to compare the predictions of the AERMOD model to the monitoring data. At present, the information provided suggests comparability of the annual averages might be satisfactory for two monitors but is extremely poor at the third receptor with underestimations on the order of a factor of 3 to 4. This comparison focuses on only one important feature of the data. The evaluation should be more extensive, and the distributions (e.g., cumulative distribution functions of concentrations, and direct comparisons between scatterplots of site-specific predicted and vs measured hourly averages) of the AERMOD short-term results should be compared with observations. The use of a homogeneous annual background to correct the AERMOD predictions does not correct the poor modeling of the spatial NO₂ concentrations across the area. Two approaches can be used to correct this problematic modeling result (the two approaches could be used in combination): (1) a more complete emission inventory can be used for input to AERMOD to provide a better representation of sources in the vicinity of the receptor where concentrations are significantly underestimated and/or (2) the modeling results can be calibrated to match the measurements at the monitoring sites.¹

The fact that only the resident population is treated in the exposure assessment should be explicitly mentioned and an estimate of the commuting population who may be exposed in Philadelphia County during working hours for example should be provided.

The cities for which there are sufficient data to perform a detailed analysis (similar to the Philadelphia analysis) should be identified. The Agency should consider the confidence intervals around the modeled air quality values in the cities under consideration. If the confidence intervals are extremely large, it may not be useful for EPA to expend resources to look at additional cities.

If the decision is made to estimate risk based on epidemiology studies, the REA will need to address co-pollutant issues. In particular, while there are limited data on possible correlations of NO₂ with other species such as particulate elemental carbon (EC), such possible correlations should be highlighted.

Exposure Analysis

Agency charge questions:

- 1. To what extent is the assessment, interpretation, and presentation of the initial results of the exposure analysis technically sound, clearly communicated, and appropriately characterized?**
- 2. The draft risk and exposure assessment document evaluates exposures in Philadelphia. Future drafts will also evaluate exposures in Atlanta, Detroit, Los Angeles, and Phoenix. What are the views of the panel**

¹ The simplest way consists of: (1) calculating the modeling error at each measurement site, (2) interpolating the error in-between the measurement sites using some distance-weighting factor (a standard factor is the inverse of the distance squared) to obtain a spatial map of the modeling error, and (3) creating a spatial map of concentrations by correcting the model results by the modeling error. The result is a map of concentrations that are equal to the measurements at the measurement sites and that follow the spatial gradients of the modeling results in between. Another approach is to kriging the modeling error to replace the interpolation in step 2; kriging is preferred if land use information can be incorporated in a universal kriging analysis.

regarding the appropriateness of these locations and on the approach used to select them?

- 3. Do Panel members have comments on the appropriateness and/or relevance of the populations evaluated in the exposure assessment?**
- 4. To what extent are the approaches taken to model stationary sources and mobile sources technically sound and clearly communicated?**
- 5. Human exposures are modeled using APEX to simulate the movement of individuals through different microenvironments. Do Panel members have comments on the microenvironments modeled?**
- 6. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability?**

Response: CASAC commends EPA on the work done thus far in the area of exposure assessment in support of risk assessment. Some aspects of the assessment, interpretation, and presentation of the initial results are to a large extent considered technically sound, clearly communicated, and appropriately characterized. The choice of cities and of populations is viewed as reasonable and appropriate. CASAC highlights several areas in which the document could be further improved. CASAC requests more text early-on to clarify the context, rationale and objectives for the analytical approaches taken, and to emphasize the relationships among, and the strengths/limitations of, the ambient and on-road exposure analyses.

There is need for more thorough evaluation of the AERMOD predictions in relation to monitoring data. This model evaluation is central to the larger question of how far the exposure modeling approach should be taken (e.g., other cities) given current uncertainties, something EPA should address at the end of Chapter 7. It is particularly important that peak hourly predictions be compared with corresponding monitoring data (e.g., using scatterplots). Features such as vertical gradients and site-to-site variations are important. A constant additive calibration factor is not adequate. The model must be assessed with respect to its variability and the tails of the distribution in order to provide appropriate justification that the AERMOD model is capturing the true heterogeneity in the ambient concentration data.

CASAC is concerned that the benchmarks chosen for analysis may not be protective of the most sensitive population subgroups – e.g., those not included for ethical reasons in chamber studies. CASAC is generally supportive of the roll-back and roll-up methods, but highlights the need for thorough accounting of indoor/outdoor relationships and cautions that these approaches may be difficult for the public to grasp. The APEX modeling is generally seen as a useful exercise, though some specific improvements are recommended. These include more realistic, e.g., log-normal, distributions on several of the input parameters (rather than uniform distributions limited to the range of available data, as used in the current analysis), appropriate justification that the AERMOD model is capturing the true heterogeneity in the ambient concentration data, and discussion of the potential co-linearities of several input parameters which may result in un-accounted for elevations in NO₂ exposures and responses among inner city residents. The APEX model plays a central role in the exposure assessment and evaluation of components of this model (or reference to a previous evaluation) is necessary. The individual comments of the panelists (Enclosure B) provide detailed additional suggestions for revising the

implementation of the APEX model for the NO_x exposure assessment. Finally, CASAC encourages EPA to take the uncertainty analyses and discussion further with the goal of conveying overall uncertainties in the final numbers presented, as well as providing insights into the relative magnitude of different uncertainty sources.

Characterization of Health Risks

Agency charge questions:

- 1. What are the views of the Panel on the overall characterization of the health evidence for NO₂? Is the presentation clear and appropriately balanced?**
- 2. The characterization of health risks focuses on potential health benchmark values identified from the experimental NO₂ human exposure literature on airways responsiveness. What are the views of the Panel on using potential health benchmarks from this literature to characterize health risks?**
- 3. Do panel members have comments on the range of potential health effects benchmark values chosen to characterize risks associated with 1-hour NO₂ exposures?**
- 4. To what extent is the assessment, interpretation, and presentation of initial health risk results technically sound, clearly communicated, and appropriately characterized?**
- 5. While the epidemiology literature will be considered in developing the Agency's policy assessment as part of an evidence-based evaluation of potential alternative standards, staff have judged that it is not appropriate to use the available NO₂ epidemiological studies as the basis for a quantitative risk assessment in this review. Do panel members have comments on this judgment and/or on the rationale presented to support it?**

There is broad consensus that three areas in the health risk characterization need considerable revision and amplification. First, the overall scientific evidence for health risk from longer-term exposures is more compelling than its characterization in the current draft. Specifically, the epidemiological data provide consistency on longer-term exposures and should be given greater emphasis (even though unequivocal interpretation is somewhat limited by the difficulty of separating the effects of NO₂ from other co-pollutants in ambient air mixtures). Health endpoints other than airways hyper-responsiveness (including respiratory illness, emergency room visits, hospital admissions, and lung function growth) should be included in the risk assessment determination. Expanded discussion of the factors that affect coherence between results from animal toxicology, clinical exposures, and epidemiological approaches is also needed. While acknowledging the methodological challenges, the exclusion of epidemiologic exposure/response relationships from the risk assessment is viewed by many as a serious shortcoming of the current draft.

Secondly, the document needs to define and distinguish “susceptible” and “vulnerable” populations, and to expand the list of potential at-risk populations under consideration. Additional at-risk groups may include those having chronic diseases or conditions other than asthma (such as obesity, cardiovascular, chronic obstructive

pulmonary disorder, and diabetes), those more vulnerable through exposure near roadways (including residents, schoolchildren, and commuters) and other categories that might lead to increased risk (including genetic pre-disposition, lower social-economic status, and smokers). A rationale should be provided for each of the identified at-risk groups.

Finally, the health evidence presented on acute indicators of risk supports consideration of a short-term (less than or equal to a 24-hour) NO₂ standard. Discussion about the form and level of such a standard should be guided by understanding of the temporal dynamics of biological responses, particularly those mediated by oxidative stress. Useful data to inform this discussion is available from both domestic and foreign studies, as well as from both outdoor and indoor research, and each of these four resource areas should be utilized.

In closing, the CASAC was pleased to review this first draft of the *Risk and Exposure Assessment* for the primary NO_x review. We look forward to reviewing the second draft of this document in September 2008, and to continuing to advise you as you complete your assessment of the NO_x primary standard.

Sincerely,

/Signed/

Dr. Rogene Henderson, Chair
Clean Air Scientific Advisory Committee

Enclosures

Enclosure A: Roster of CASAC Oxides of Nitrogen Primary NAAQS Review Panel

Enclosure B: Compilation of Individual Panel Member Comments on EPA's Risk and Exposure Assessment to Support the Review of the NO₂ Primary National Ambient Air Quality Standard: First Draft

Enclosure A
U.S. Environmental Protection Agency
Clean Air Scientific Advisory Committee
Oxides of Nitrogen Primary NAAQS Review Panel

CHAIR

Dr. Rogene Henderson, Scientist Emeritus, Lovelace Respiratory Research Institute, Albuquerque, NM

CASAC MEMBERS

Dr. Ellis B. Cowling, * University Distinguished Professor At-Large, Emeritus, Colleges of Natural Resources and Agriculture and Life Sciences, North Carolina State University, Raleigh, NC

Dr. James Crapo, Professor of Medicine, Department of Medicine, National Jewish Medical and Research Center, Denver, CO

Dr. Douglas Crawford-Brown, Professor and Director, Department of Environmental Sciences and Engineering, Carolina Environmental Program, University of North Carolina at Chapel Hill, Chapel Hill, NC

Dr. Donna Kenski, Data Analyst, Lake Michigan Air Directors Consortium, Des Plaines, IL

Dr. Armistead (Ted) Russell, Professor, Department of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA

Dr. Jonathan M. Samet, Professor and Chair of the Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

CONSULTANTS

Mr. Ed Avol, Professor, Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA

Dr. John R. Balmes, Professor, Department of Medicine, Division of Occupational and Environmental Medicine, University of California, San Francisco, CA

Dr. Terry Gordon, Professor, Environmental Medicine, NYU School of Medicine, Tuxedo, NY

Dr. Dale Hattis, Research Professor, Center for Technology, Environment, and Development, George Perkins Marsh Institute, Clark University, Worcester, MA

Dr. Patrick Kinney, Associate Professor, Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY

*Unable to participate in the May 1-2, 2008 CASAC Panel Meeting

Dr. Steven Kleeberger, Professor, Lab Chief, Laboratory of Respiratory Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC

Dr. Timothy V. Larson, Professor, Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA

Dr. Kent Pinkerton, Professor, Regents of the University of California, Center for Health and the Environment, University of California, Davis, CA

Dr. Edward Postlethwait, Professor and Chair, Department of Environmental Health Sciences, School of Public Health, University of Alabama at Birmingham, Birmingham, AL

Dr. Richard Schlesinger, Associate Dean, Department of Biology, Dyson College, Pace University, New York, NY

Dr. Christian Seigneur, Vice President, Atmospheric & Environmental Research, Inc., San Ramon, CA

Dr. Elizabeth A. (Lianne) Sheppard, Research Professor, Biostatistics and Environmental & Occupational Health Sciences, Public Health and Community Medicine, University of Washington, Seattle, WA

Dr. Frank Speizer, Edward Kass Professor of Medicine, Channing Laboratory, Harvard Medical School, Boston, MA

Dr. George Thurston, Professor, Environmental Medicine, NYU School of Medicine, New York University, Tuxedo, NY

Dr. James Ultman, Professor, Chemical Engineering, Bioengineering Program, Pennsylvania State University, University Park, PA

Dr. Ronald Wyzga, Technical Executive, Air Quality Health and Risk, Electric Power Research Institute, Palo Alto, CA

SCIENCE ADVISORY BOARD STAFF

Dr. Angela Nugent, Designated Federal Officer, 1200 Pennsylvania Avenue, NW 1400F, Washington, DC, Phone: 202-343-9981, Fax: 202-233-0643, (nugent.angela@epa.gov)

Enclosure B: Compilation of Individual Panel Member Comments on EPA’s Risk and Exposure Assessment to Support the Review of the NO₂ Primary National Ambient Air Quality Standard: First Draft

Comments from Professor Ed Avol.....	10
Comments from Dr. John Balmes.....	12
Comments from Dr. Douglas Crawford-Brown.....	14
Comments from Dr. Terry Gordon.....	19
Comments from Dr. Dale Hattis.....	21
Comments from Dr. Donna Kenski.....	25
Comments from Dr. Patrick Kinney.....	27
Comments from Dr. Steven Kleeberger.....	29
Comments from Dr. Timothy Larson.....	31
Comments from Dr. Kent Pinkerton.....	35
Comments from Dr. Armistead Russell.....	37
Comments from Dr. Jonathan Samet.....	40
Comments from Dr. Richard Schlesinger.....	42
Comments from Dr. Christian Seigneur.....	43
Comments from Dr. Elizabeth “Lianne” Sheppard.....	45
Comments from Dr. Frank Speizer.....	50
Comments from Dr. George Thurston.....	51
Comments from Dr. James Ultman.....	53
Comments from Dr. Ronald Wyzga.....	54

Comments from Professor Ed Avol

Air Quality Information and Analyses (Chapters 2,5, 6)

P65, Table 16 – This is an informative and useful table, in that it identifies potential sources of error in the ensuing assessments and provides some insights into Agency weighting of the component categories. The utility of the table is somewhat compromised by magnitude assessments of “minimal” or “moderate”, which infer no absolute quantity or range of effect, but the listing of sources and types is appreciated.

Exposure Analysis (Chapters 5, 7)

Much of this detail about how APEX and AERMOD and CHAD actually functions seems more appropriate for an appendix, rather than the main body of the report.

P73, lines 3-7 (selection of upper-air station locations for the respective cities to be modeled) – What implications do significant distances between the city being modeled and the upper-air station location have -- Philadelphia is using Washington Dulles data, Los Angeles is using San Diego data, Phoenix is using Tucson data...is this appropriate? Should some comment be made about this?

P78, Table 19 – From a Los Angeles perspective, these AADT figures look low – are maximum freeway values (in one direction) really only ~68,000 vehicles?

P80, Table 20 – Coming from Los Angeles, it's difficult to believe that average speed on freeways can actually be 62 to 66 miles per hour! This would seem to me to be a high estimate of actual traffic flow – what about inclement weather (snow, rain, etc)? What is the time period over which the average is determined?

P81 through p88, Other Emission parameters – Appropriately, roadway traffic, stationary sources, fugitive, and airport emissions are considered in the NOx inventory...but what about off-road activities (construction, yard equipment, etc)? What about port or dockside activities

(propulsion and auxiliary engine operations from ships, harbor craft activities, recreational boating)? What about rail?

P99, Table 29 – I assume air conditioning prevalence estimates are quite high for Phoenix, but no data is currently provided...is it not available, in process, or unreliable?

P102, line 19 – what is the basis for the “exactly one hour” stipulation for cooking events? I would think that most cooking events (those involving a stove) require more than an hour), but would have some diminished in-house emissions compared to stove-top cooking with open flames, which would result in much higher in-house emissions (but maybe not be quite so long)...

Characterization of Health Risks (Chapters 3,4, Sections 6.3, 7.8, 7.9)

P12-13, Chapter 3 – At-Risk Populations: The document identifies three sub-categories for discussion – disease/illness, age, and proximity to roadways – but others were discussed and “accepted” in the ISA. What about genetic susceptibility? What about a pre-natal component of the “age” sub-category? What about those in confined-space working conditions (such as parking garages)?

Chapter 4 does a nice job of summarizing the identified literature.

Chapter Sections 7.8 and 7.9 are detailed and involved, and are strongly dependent on the input assumptions presented earlier in the chapter (see questions above [in Exposure Analysis comments] regarding some of these assumptions).

Comments from Dr. John Balmes

General Comments

Characterization of Health Risks (Chapters 3 and 4 and Sections 6.3, 7.8, and 7.9)

1. What are the views of the Panel on the overall characterization of the health evidence for NO₂? Is this presentation clear and appropriately balanced?

I find the overall characterization of the health evidence concerning ambient NO₂ exposures to be well presented and to reflect the presentation and discussion of this evidence in the draft ISA. The specific concern that I have with the way the evidence is characterized in the ISA is also germane for this document. That is, I find the evidence that long-term exposure to NO₂ affects growth of lung function in children to be more compelling than the staff judgment.

2. The characterization of health risks focuses on potential health benchmark values identified from the experimental NO₂ human exposure literature on airways responsiveness. What are the views of the Panel on using potential health benchmarks from this literature to characterize health risks?

While I understand the why the staff decided to use the experimental data on airways responsiveness in asthmatic adults to identify potential health benchmark values to characterize risk from exposure to ambient NO₂, I would have preferred to see asthma exacerbation data (hospital admissions, emergency department admissions, asthma symptoms) used. These endpoints are easier for members of the policy audience to understand.

3. Do panel members have comments on the range of potential benchmark values chosen to characterize risks associated with 1-hour NO₂ exposures?

I find the range of potential benchmark values to be reasonable.

4. To what extent is the assessment, interpretation, and presentation of initial health risk results technically sound, clearly communicated, and appropriately characterized?

With the strong caveat that I would have preferred the asthma morbidity endpoints associated with NO₂ exposure in epidemiological studies to be used as potential health benchmarks rather than airways responsiveness, I find that the assessment, interpretation, and presentation of the initial health risks to be done well.

5. While the epidemiology literature will be considered in developing the Agency's policy assessment as part of an evidence-based evaluation of potential alternate standards, staff have judged that it is not appropriate to use the available NO₂ epidemiological studies as the basis for a quantitative risk assessment in this review. Do panel members have comments on this judgment and/or the rationale presented to support it?

While I understand the rationale for the staff judgment presented in section 4.2.3.3, I am not persuaded that the judgment is necessarily the correct one. Although many of the epidemiological studies on the effects of short-term exposure to NO₂ have been conducted outside of the United States, in my view the results of the relatively small number of U.S. studies is consistent with those of the non-U.S. studies so that the entire body of epidemiological literature could be used to develop concentration-response relationships. I am also not as concerned as staff about trying to identify an independent effect of NO₂ from the combined effect of the traffic-related pollutant mixture because NO₂ appears to be the best single pollutant marker of this mixture. Traffic-related pollution has been strongly associated with health effects and needs to be better controlled. The current single-pollutant regulatory focus of the agency does not incorporate the research data which suggest that the total oxidant pollution burden in the ambient air is responsible for health effects.

Specific Comments

p. 16, lines 17-19 The term, airways responsiveness, usually refers only to lung function responses rather than to inflammatory responses. Therefore, I would revise this sentence as follows: “*Airway responses* can be measured...”

p. 106, line 6 should be “...dispersion modeled concentrations were not rolled-up...”

p. 108, line 17 should be “...a greater number of annual average concentrations *was* estimated...”

Comments from Dr. Douglas Crawford-Brown

These comments focus on Chapters 5 and 7 of the Risk and Exposure Draft and the associated sections of the Technical Support Document, referring to other chapters only as they are needed to make points raised in these two chapters. This first draft focuses solely on risks and exposures associated with the current ambient levels, and with exposures that would occur if the current NAAQS is met throughout the country. It does not address the impacts of potential changes in the NAAQS, which was a bit surprising at first reading. I believe it would have been better to just develop the full assessment. However, in doing it in the current order, I suppose this provides an opportunity for the CASAC to comment on the methodology first before the full assessment is conducted for all scenarios. So these review comments are provided in this vein.

On a very broad issue, I compared the conclusions in the early chapters to those in the ISA, and the authors have been faithful to that earlier document. The same health effects are considered, and the same exposure durations are considered. The current document also uses the sensitive subpopulations recommended by the ISA. It also places the same caveats (strengths and limitations) on the ability to estimate personal exposures. The one exception I note is that on Page 28, the authors of the current document conclude that ambient exposures are a reasonable surrogate for personal exposures. I am not sure the ISA fully supports this conclusion, or at least does not state it so directly. The ISA left the impression that there are a number of limitations in the use of the ambient exposure data. These limitations would not be so important in applying the results of epidemiological studies, since these are based on ambient exposure measures as well. But the difference can be important if clinical studies are used to estimate relationships between exposures and effects. Still, there is no way to improve upon the approach used in the current document, and so this issue is more of a scientific than a risk assessment and policy one.

In previous reviews of NAAQS assessments, I have generally approved the proportional roll-up or roll-down methods based on current maximum concentration at a specific site. I support, therefore, the use of this method in the current document. The authors could improve the document, however, by noting that this process implicitly assumes that regulated sources and non-regulated sources are equally affected by any change in the NAAQS, or that the regulated sources will dominate the exposures.

I agree that the adjustment of the benchmarks produces the same result mathematically. But it makes no sense scientifically, and is likely to be attacked as such. The savings in processing time don't appear to me sufficient to justify a method that people will fail to understand as mathematically equivalent, and will make it appear that the EPA staff is willing to make calculations based on an assumption that effects occur at levels below the benchmarks. This doesn't seem to me to be a politically wise strategy, especially given modern computing times.

I support what is essentially a hazard quotient or exposure margin approach in the assessment. I can see no alternative to this given the lack of a reliable exposure-response curve on which to perform more detailed assessments. The one issue I would raise here is that the hazard quotient approach usually has a margin of safety built in through uncertainty factors, and the current assessment does not appear to have this margin built in. Perhaps it is buried inside the benchmarks, but I can't find that stated directly. The staff should consider how to address this issue in the methodology. This can perhaps be done by simply noting that the studies were in humans, and many were in sensitive subpopulations, and so uncertainty factors are not needed (or are set to 1).

Chapter 5 does not adequately describe what the authors mean by a two-step approach. It is clear from the writing that the first step uses only the ambient monitors and the second involves corrections for personal factors (activity patterns, etc), but it is not clear from the writing whether the first step is simply the input into the second or whether it is to be a competing analysis to the second. I assumed at first the former is the case, but the text doesn't make it clear and there is even wording at the beginning of Chapter 5 to suggest otherwise. And then the two different sets of results in Chapters 6 and 7 make it seem I was wrong in this assumption of Step 1 being the input to Step 2. Each approach has its limitations, as the first step fails to include personal differences but the second may be introducing personal differences that are already reflected in benchmarks. This latter issue is always important when epidemiological results are used, as the exposure categories usually are based on ambient results but the risk coefficients have buried within them the interpersonal variations in the ratio of personal exposure to ambient levels.

Having said this, I fully support the use of APEX and CHAD for the purpose of performing these stochastic calculations IF inter-subject variability of exposure is appropriate to estimate. These models contain assumptions that are quite routine in EPA assessments and have found application in a wide range of settings. They have been fully vetted for the kinds of assessments performed here. The one issue I would raise is that there remains the problematic relationship between ambient levels as measured at monitors and ambient levels at or near the points of exposure for populations. This is particularly important in Step 4 on Page 68. I suppose there is not much that can be done about that issue, because the monitors are located where they are and can't be changed for the purposes of this assessment. But I would like to see a slightly better description of the implications of this problem when APEX is run. And in any event, as I comment later, it is evident that the monitoring results are not in fact input to the calculations of intersubject variability in Step 2.

The site selections were good given the 90th percentile rule specified. So the assessment results should characterize the upper bounds of exposure in the more heavily polluted communities. I doubt it will capture exposures at small geographic areas that might have multiple sources of NO_x, unless monitors are already located there (I am not convinced they are). But given these limitations, the sites chosen seem to me reasonable.

As my expertise does not extend to air quality modeling of the type performed here, I can't comment on the adequacy of AERMOD for these purposes. It is a modeling package that has been extensively used in past EPA assessments, and so I will assume here that it has been vetted. I don't, however, understand how the assessors have combined the air monitoring data and the model results. I had thought the air monitoring data were being used to establish ambient levels, but this must not be the case since AERMOD is being used to estimate exposures based on emissions inventories (and since Chapter 6 results are apart from those in Chapter 7). The early sections of this document would be improved by making it clear how the monitoring and modeling results are to be combined. It appears that I may have been wrong in Chapter 5 in assuming that the national monitoring results were the inputs to the second step of the assessment (the step that generates inter-subject variability in exposure). If I am confused, others might be as well.

Assuming the air modelling can be performed adequately (and again, I will leave it to other CASAC members to comment on this in a more informed way), then the subsequent steps are reasonable. The development of the longitudinal activity sequences is a sophisticated piece of work, being state-of-the-science. The stochastic sampling methodology is reasonable and employed commonly at the EPA. The assumptions going into the sampling are adequately described. The microenvironments are both the correct ones to model given current data and well executed in the assessment steps. The equation on Page 105 is the correct one to use for calculating time-averaged exposure for the period considered.

I found the characterization of results at the end of Chapter 7 informative and simple to follow. They walk the reader through the relevant findings. While I found the results showing the contribution of different microenvironments interesting, I am not sure how it will be used in any decisions on a NAAQS. I do imagine it might be useful in determining WHICH microenvironments should be the focus of attention in changing the relationship between ambient levels and micro-environmental levels, but the NAAQS will in the end be based on ambient levels. Perhaps the authors could just place in the document a few comments on why these results are of interest.

The analysis of repeat exposures (around Page 120) falls into the same category of results that are quite interesting scientifically, but where the policy implications are not clear. My experience is that the EPA tends to treat one individual with N episodes the same as N individuals each with one episode. Again, just some clarification on the significance of this analysis would be useful

I found it difficult to follow the variability and uncertainty analyses. Part of the problem is that the discussion moves pretty fluidly between variability and uncertainty considerations, and so I was never completely clear what was being considered as variability and what was being considered as uncertainty. And it seems to me that the uncertainty part just doesn't even touch on important sources of uncertainty or provide a good description of how well predicted effects results are expected to compare with measured effects (if the latter were available). Instead, the uncertainty assessment focuses primarily on the contribution of a few key elements of the

assessments to the uncertainty. I expected to see some statements, even if qualitative, about the uncertainty in the various risk results (e.g. uncertainty in number of people above a benchmark, percent of asthmatics experiencing a high exposure day). This aspect can be greatly improved.

I end with a comment I have made in other settings of CASAC. The modelling performed here is impressive and represents state-of-the-science. But I worry that it may be too elaborate for the purposes of establishing a NAAQS, particularly if a party were to try to delay a NAAQS by attacking one assumption at a time. There are many, many assumptions built into the assessment. I had thought from Chapter 5 that the monitoring results were going to play a more central role, and that the personal exposure and risk results would apply some kind of post-processing correction factor to the monitoring results. But it is now evident that this is not the case, and that Chapter 6 stands quite alone from Chapter 7. We will need to discuss that in more detail at the CASAC meeting. Perhaps the authors might find a way to compare the two results more systematically and see how well the mean exposures and risks compare for areas that are common in Chapters 6 and 7.

Comments from Dr. Terry Gordon

Charge Questions:

1. The presentation on the overall characterization of the health effects is clear and well-balanced.
2. Airway hyperresponsiveness (BTW, it's not usually written as 'airways') is appropriate as one benchmark. The choice of using this health endpoint solely is somewhat controversial. This benchmark effect, seen in asthmatics after short-term exposure, was not seen in every clinical study although it appeared to be seen consistently in resting test subjects and not in exercising subjects. So, while it is a suitable benchmark, it has some weak points as does the epidemiology literature in separating the health effects of NO_x from co-pollutants.
3. The choice of a 1 hour benchmark-related health endpoint is logical given the database of effects in clinical studies.
4. The assessment, interpretation, and presentation of the health risk results are satisfactory and, for the most part, clearly presented. There is an opportunity to make small changes and polish the presentation in the next draft.

Minor Comments:

Page 12, line 20 – The text states an association between NO₂ and cardiac effects. The statement is somewhat misleading given the text on page 25, line 11 and page 26, line 17 which point to 'inadequate' evidence.

Page 14, lines 3-19 – The use of criteria/decisive factors to delineate findings into categories is a good approach. While it may not be optimal (some CASAC members will likely suggest some tweaking), it is good basis for decision-making in this assessment.

Page 15, line 7 – Unclear, should 'as high as' be 'as low as'?

Page 16, line 10 – Add 'specific' before responsiveness.

Page 16, lines 17-19 – Airway responsiveness is assessed by pulmonary function changes and does not typically refer to inflammation.

Page 19 – The table and the Annex do not list the study by Orehek et al, 1976 which found airway hyperresponsiveness in human subjects exposed to 0.1 ppm NO₂ (Orehek J, Massari JP, Gayrard P, Grimaud C, Charpin J. Effect of short-term, low-level nitrogen dioxide exposure on bronchial sensitivity of asthmatic patients. J Clin Invest. 1976 Feb;57(2):301-7.).

Page 42, line 22 – Is the higher potential for Detroit or overall?

Page 45 – In these tables, does there need to be a column of minimum exceedances. Isn't this always zero unless all the data for that monitor is over the benchmark level?

Page 106, line 6 – Typo: concentration(s)

Page 117 – I wonder about the adequacy of the model if the time spent outdoors in a parking lot almost equals the time spend inside the residence.

Page 123, line 11 – Typo: should refer to Figure 18

Page 132, line 12 – Has CARB been defined? Should there be a reference for this?

5. Omitting the epidemiology data on respiratory health effects from the quantitative risk analysis may be a rash decision. The epidemiology data, although always lacking in terms of proving causal relationships, are strong for some respiratory endpoints. These endpoints, although confounded by co-exposure to other pollutants, are consistent and backed up by the clinical and toxicology data. Therefore, the use of epidemiology data regarding respiratory effects should be considered as a strong candidate for additional quantitative risk analysis.

Questions on the Exposure Analysis Portion (Chapters 5 and 7) of the Exposure and Risk Analysis Document

1. To what extent is the assessment, interpretation and presentation of the initial results of the exposure analysis technically sound, clearly communicated and appropriately characterized?

The authors of the document have done a great deal of work in modeling air quality in Philadelphia and how this might change under a number of roll-up scenarios. Unfortunately the use of the available information to estimate changes in exposures to the general population and specific sensitive groups under those scenarios has several serious deficiencies that must be corrected before they can be used in assessing options for a revised NO_x NAAQS.

With the additional explanation provided at the meeting, the roll-down approach for calculating equivalent exceedances for the stated benchmarks appears to be capable of providing reasonable estimates of the effects of policy-related changes in outdoor ambient levels on total exposures. However it is critical that the upper tail of the hourly distributions exposures from indoor sources are modeled reasonably. Figure 7 indicates that indoor sources overall contribute about a third to annual average NO₂ concentrations. Figure 8 clearly indicates that indoor sources contribute appreciably to maximum exposures for individuals for the median and upper percentiles, although the variability is clearly greater for outdoor-source exposure estimates in the current model. This may in part be the result of artificial truncation of the uniform distributions used for NO₂ indoor source contributions and removal rates (see further discussion below). Figure 11 presents the results in more dramatic form—indicating a fivefold difference in the estimated number of asthmatics exposed above 200 ppb at least once per year.

An important problem in the current analysis relates to the adjustment of the source + dispersion model predictions to correspond to the observed data from air quality monitors. It is good that the authors made such a comparative reality-check. However from the comparison in Table 26 (p. 91) it appears that the monitors report much more consistency in the annual mean concentrations among different places than the model predicts. This suggests that the models are underestimating NO₂ concentrations attributable to background/long range transport in comparison to local sources. However the comparison is based only on long term averages. It is essential to compare predicted and observed hourly time distributions, as these are the critical inputs for the health risk analysis in its current form. It is not at all clear that addition of a uniform number for the arithmetic mean for each receptor will result in an accurate correction of the hourly concentration distribution. Because hourly concentrations are influenced by short term meteorological data, it is possible that a multiplicative correction approach might more accurately reflect changes needed to the modeled distributions of exposures for shorter averaging times. Finally, it is not clear how the corrections for the receptors shown are applied to the diverse geographic locations of all the receptors in Philadelphia. Table 26 shows the derivation of corrections for only 3 monitoring sites. These are apparently the only sites with available monitoring data in the study area. However it is still important for the authors to discuss in some

detail how the data for these sites is applied to the locations of the thousands of specific receptors that are included in the APEX modeling effort

A second issue relates to this same correction approach in another way. The model predictions are appropriately designed to represent concentrations at the assumed height of human receptors, 1.8 meters. However if they are adjusted to observed data collected at greater heights, then because NO₂ levels are thought to decline with elevation, it is likely that the corrections derived are smaller than they should be. In the uncertainty section on page 53 of the TSD there is only one small set of statements, not even a whole paragraph, devoted to this likely bias:

“Also, negative vertical gradients exist for monitors (2.5 times higher at 4 meter vs. 15 meter vertical siting (draft ISA, section 2.5.3.3), thus monitors positioned on rooftops may underestimate exposures. Only 7 of the 177⁹ monitors in the named locations contained monitoring heights of 15 meters or greater, with nearly 60% at 4 meters or less height, and 80% at 5 meters or less in height. Not accounting for this potential vertical gradient in NO₂ concentrations may generate underestimates of exceedances for some site-years, however the overall impact of inferences made for the locations included in this assessment is likely minimal since most monitors sited at less than 4-5 meters in vertical height.”

Instead of this essential dismissal of the problem, at least one panelist recommends analyzing it in the following way to estimate the likely extent of the bias.

Model the decline in concentrations with altitude, given available data. Preliminary calculations by one panelist used exponential and Gaussian models. However the detailed air modeling results derived by the study team itself may make explicit predictions about the expected decline in NO₂ concentrations with elevation.

Multiply the fraction of monitors in each height interval by the approximate amount of multiplicative bias relative to a receptor at about 2 meters. Early calculations by a panelist yield indicate that something like a 17-35% correction is needed to convert airborne monitored concentrations to equivalent 2 meter concentrations. The specific cases of the three available receptors in Philadelphia should be analyzed in detail.

A third issue is the representation of a few key sources of variability in the APEX exposure modeling:

Air exchange distributions contingent on temperature and presence or absence of air conditioning. Overall the panel does not have any objection to the idea of using lognormal distributions with very broad limits (.1 and 10 air changes/hr). However the detailed results seem to show different patterns with temperature arbitrarily blocked into a few ranges. There does not appear to be any great consistency or overall theory for this analysis. A better description of the data as a whole might be produced by a more extensive regression study using temperature or some transform of temperature as a continuous variable and either fixed-effect or mixed effects modeling of differences among cities and for the air conditioner presence variable.

NO2 removal rate distribution--p. 101. At least one panelist expressed an objection to the narrow fixed limits used for the removal rate distribution based on six values from Spicer et al (1993). The abstract to the Spicer paper makes it clear that all six observations were made in a single house, and that there are additional complications from the presence of HONO, an apparently longer-lived NO_x species:

p. 101—The same panelist also objected to the fixed limits used for the removal rate distribution based on six values from Spicer et al (1993). The abstract reads.

Transformations, lifetimes, and sources of NO₂, HONO, and HNO₃ in indoor environments.

Spicer CW, Kenny DV, Ward GF, Billick IH.

Air Waste. 1993 Nov;43(11):1479-85.

Battelle, Columbus, OH 43201-2693.

Recent research has demonstrated that nitrogen oxides are transformed to nitrogen acids in indoor environments, and that significant concentrations of nitrous acid are present in indoor air. The purpose of the study reported in this paper has been to investigate the sources, chemical transformations and lifetimes of nitrogen oxides and nitrogen acids under the conditions existing in buildings. An unoccupied single family residence was instrumented for monitoring of NO, NO₂, NO_y, HONO, HNO₃, CO, temperature, relative humidity, and air exchange rate. For some experiments, NO₂ and HONO were injected into the house to determine their removal rates and lifetimes. Other experiments investigated the emissions and transformations of nitrogen species from unvented natural gas appliances. We determined that HONO is formed by both direct emissions from combustion processes and reaction of NO₂ with surfaces present indoors. Equilibrium considerations influence the relative contributions of these two sources to the indoor burden of HONO. We determined that the lifetimes of trace nitrogen species varied in the order NO approximately HONO > NO₂ > HNO₃. The lifetimes with respect to reactive processes are on the order of hours for NO and HONO, about an hour for NO₂, and 30 minutes or less for HNO₃. The rapid removal of NO₂ and long lifetime of HONO suggest that HONO may represent a significant fraction of the oxidized nitrogen burden in indoor air.

The uniform distribution with its fixed boundaries (0% probability assumed for values outside of the defined limits) is particularly inappropriate when the data are limited, as in this case. Use of the uniform distribution artificially reduces the likelihood of more extreme values of the modeled parameter than happen to be present in the limited available data. This in turn limits the model-predicted variability of NO₂ concentrations, which critically determines the number of exceedances of the high hourly NO₂ levels that are the focus of the risk assessment modeling. It would likely be far better to use a lognormal here as an initial hypothesis, but in the light of the fact that different houses with different internal materials might well destroy NO₂ at different rates, expert

judgment might well be needed to expand the likely distribution beyond what can be derived from a simple data fit.

The same panelist also strongly objected to the use of uniform distribution of concentrations of NO₂ from use of gas stoves (p. 101). The very breadth of the bounds derived (4 – 188) ppb argues against a uniform distribution and in favor of something more skewed, such as a lognormal. The lognormal guarantees a positive contribution, and doesn't have the unfortunate property of implying zero chance that the indoor contribution will be above the derived maximum. Moreover, if a mass balance approach is being used to model indoor NO₂, then the input per cooking event should be in terms of mass units of NO₂, not concentration. Concentration will depend on house- and temperature specific factors such as air exchange rates, NO₂ removal rates and residual contributions from HONO, among other things. Because these observations were from a single house in California, there must be extra allowance for variability and uncertainty in these estimates that must clearly extend beyond the mass equivalent of the concentration range quoted.

Finally the assumption that all cooking events contributing to indoor NO₂ last exactly one hour also artificially limits the variability in NO₂ inputs and therefore exposures represented in the model.

Comments from Dr. Donna Kenski

To what extent are air quality characterizations and analyses technically sound, clearly communicated, appropriately characterized, and relevant to the review of the primary NO₂ NAAQS?

To what extent are the properties of ambient NO₂ appropriately characterized, including ambient levels, spatial and temporal patterns, and relationships between ambient NO₂ and human exposure?

These two questions actually seem more suited to the ISA than to the REA, but generally the air quality representation in the REA was fine. It seems that great care was taken in screening and cleaning the NO₂ data for use in this assessment, and that process was described thoroughly. Section 2 was very brief, but adequate given that it was comprehensively discussed in the ISA. My concerns with the air quality characterization have mostly to do with the roll-up and the roadway treatment, described below.

In order to simulate just meeting the current standard, we have rolled up NO₂ air quality levels. To what extent is the approach taken technically sound, clearly communicated, and appropriately characterized? Do Panel members have comments on the relevance of this procedure for reviewing the primary NO₂ NAAQS?

I'm inclined to think this approach is satisfactory, but I have a nagging doubt that in rolling up air quality we have somehow inflated the role of outdoor sources and underestimated the impact of indoor sources. I would like the document to convince me otherwise. Discussion of why this may or may not be the case would be welcome.

We have evaluated air quality in a number of individual locations throughout the United States. What are the views of the panel regarding the appropriateness of these locations and on the approach used to select them?

I thought the evaluation and selection of specific locations was well developed and entirely appropriate.

Because of the impact of mobile sources on ambient NO₂, we have estimated on-road NO₂ concentrations. To what extent is the approach taken technically sound, clearly communicated, and appropriately characterized? Do Panel members have comments on the relevance of this procedure for reviewing the primary NO₂ NAAQS?

I don't quite see the utility of this particular on-road estimation method (Sec. 6.2.3). I was happy enough with the relationship described in Eq 2, and with the model for predicting m as described in the TSD, although it would also be nice to see values for k described here. But to generate on-road concentrations for all monitors randomly, without regard for where the monitors are, roadway size or type or number of vehicles per day, and then make the conclusion that roadways with high vehicle densities are likely better represented by estimates at the upper tails, seems like a lot of work to reach an obvious conclusion. I'm not sure the numbers are meaningful, just because there are a lot of them. Perhaps this section just needs to communicate more clearly the

purpose of these simulated concentrations. Or explain why a model that incorporates information about traffic densities or roadway size wasn't used to generate this distribution of concentrations?

What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability?

Sec. 6.4 needs editing for language and punctuation but was exceedingly helpful in describing the various sources of uncertainty and the magnitude of potential influences on results; likewise, Table 16 was a great summary of the information presented. I only have one slight reservation, and that is about the assumption that similar sources impact both simulation time periods equally (Sec. 6.4.3). In fact the NOx SIP call influenced utility industry emissions significantly in the more recent period. It is not apparent how this constitutes a 'minimal' bias, as indicated in Table 16. Perhaps an acknowledgment of some significant source changes would be warranted, or an indication of how this impact was determined to be minimal could be provided.

Other specific comments:

Please add a list of abbreviations to the front matter.

p. 78, Table 19: define CBD (central business district?)

p. 109, Fig. 5: It is hard to distinguish these lines because the symbols blur together; use colors?

p. 109, line 12: not clear "...persons estimated to contain exposures...?"

Figures 9-11 and 14-18 are made more difficult to interpret because of the unnecessary use of 3-D, which makes it much harder to judge the relative positions/heights of the bars. These would be much more effective as simple 2-D bar graphs. Use color or patterns to distinguish between groups.

Comments from Dr. Patrick Kinney

1. Chapters 5 and 7

General responses to charge questions:

Overall I commend EPA staff on this initial draft. To a large extent, it is technically sound, well-written, and interpreted. The choice of study locations are well-justified and appropriate. The selection criteria are clearly stated and sound.

The decision to focus mainly on asthmatics seems reasonable given our current understanding of NO₂ health risks. When it comes time to estimate exposures in chapter 7, however, I questioned whether census-block/track-specific asthma rates were used to estimate the population at risk, or was city-wide asthma prevalence used instead. This touches upon a principal question/critique at this stage, which is that the analysis needs to either analyze or else discuss the implications of not analyzing the differential risks that may arise for inner city residents who 1) may have higher than average asthma prevalence, 2) may have higher than average exposures to traffic emissions (and which may not have been accounted for if only “major roadways” were included in the source term), 3) may be more likely to spend time outdoors and on-foot, 4) may be less likely to use air conditioning and thus receive higher ambient contribution to indoor levels. The APEX model is very impressive in its scope but it is important to recognize that input parameters are not necessarily independent of one another, and may instead be somewhat co-linear on economic or racial gradients.

The modeling of stationary and mobile sources seems to have been done well. I do think however that there should be discussion and/or sensitivity analyses presented on the issue of major vs. all roadways. Only major were included in the model. What proportion of roadway NO_x emissions within Philadelphia country are lost in making this choice? Forgive me if I missed this detail someplace.

The microenvironments chosen for the APEX modeling make good sense. However, it was unclear to me whether pedestrian movement along roadways was modeled explicitly. We would expect the sidewalk “microenvironment” to have higher exposures than home or work, even if in the same census block – both because of lower vertical height and proximity to roadway sources.

The assessment of uncertainty represents a very good first effort. This section is likely to grow with subsequent drafts.

2. Chapters 3 and 4 and Sections 6.3, 7.8 and 7.9

Overall, the characterization of the health knowledge base is well done. I concur with staff decision to base the analysis on human chamber study results, and utilize epidemiology results a supportive role. One issue that deserves greater emphasis, however, is the fact that human chamber studies do not capture the most sensitive tail of the population susceptibility distribution and thus inherently represent overestimates of benchmark concentrations and conversely underestimates of health impacts at a given concentration. This concept does not seem to have

been incorporated in choosing the 200-300 ppb range of health benchmarks, insofar as the document states on p. 16, line 25 that 76% of subjects responded within that range.

Specific comments throughout the document:

Page 30, line 18. In what sense are these “scenario-driven” analyses? This term doesn’t seem appropriate here. If it is appropriate, we need to understand how; add explanatory text.

This paragraph also is a good place to explain the rationale for these two approaches, what were their specific objectives, and how the two relate to one another.

Page 35, section 6.1: This section is really an overview of the methodology. Still missing is the context, rationale, and major objectives of this methodology. What is it intended to tell us about exposure and health risks? What can it do and what are its limitations?

Page 47, line 11, change section ref to 6.2.3

Page 54, line 20, check section ref.

Page 58, line 26, insert “that source” after “influence”

Page 66, section 7.1. Again, need to lead off this section with a clear and concise statement of context, rationale, and objectives for this set of work. How does it fit into the big picture of assessing risks? What are strengths and limitations with respect to the Chapter 6 approach?

Page 69 top, Does the model take into account higher NO₂ near the ground and near roadways?

Page 72, line 24. What is meant by “mandatory and significant?”

Page 91, table 26. I’m troubled with the big differences observed, even after “correction.” Probably need more reassuring explanation for the non-modelers.

Page 93, line 14. I would prefer using the mean. Zero seems quite unlikely.

Page 93, lines 15-17. This implies that commuting by sidewalk and/or bus was not accounted for, which I find problematic in the inner city.

Page 102, line 4. Inner city cooking patterns may be quite different, with longer hours spent preparing meals.

Page 119, line 25. Edit for grammar.

Comments from Dr. Steven Kleeberger

Sections 3 (At Risk Populations) and 4 (Health Effects)

As per the directive for this document, the report focuses on studies that have been published in the peer-reviewed literature, with exposure duration and concentration with reasonably acceptable ranges (i.e., those potentially experienced in indoor and outdoor (not occupational) environments). Based on this criterion, two major health effects were identified: increased AHR in asthmatics with short-term exposures and increased respiratory infections in children with long-term exposures. Also identified were subpopulations considered potentially more susceptible to the effects of NO₂ include: individuals with preexisting respiratory disease; children; elderly.

Additional Comments: question 4 – “What are the views of the Panel on the characterization of groups likely to be susceptible or vulnerable to NO₂ and the potential public health impact of NO₂ exposure?”

A distinction is made between “susceptibility” (disease- and age-mediated) and “vulnerability” (children and elderly). Susceptibility therefore appears to describe those factors that may be considered “host” or “intrinsic” while vulnerability appears to be related to an interaction between susceptibility (risk for adverse outcome based on intrinsic risk factor) and increased risk of enhanced exposure. It therefore seems that these descriptors are not mutually exclusive, and the utility of the two terms is not entirely clear.

I would like to suggest that a table or figure be included in the document to more clearly identify subgroups or subpopulations that are likely susceptible to adverse effects of exposure to NO₂. Currently, the report also focuses on three important subpopulations that are potentially susceptible to NO₂ effects, including individuals with preexisting respiratory disease (asthma), children (enhanced risk of respiratory infection), and the elderly (compromised antioxidant defense). A number of investigations of host susceptibility for adverse health effects of exposure to other air pollutants (notably PM and ozone) indicate that nutrition, obesity, genetic background, etc are important in human studies and animal models. I suggest that additional pre-existing diseases such as obesity should also be included. Clinical and animal studies have demonstrated that this disease is a risk factor for adverse response to O₃ exposure. A pre-existing condition that could be included is very low birth weight (VLBW). Prematurity has been demonstrated to be an important risk factor for respiratory virus infections and higher incidence of asthma, and it is not unreasonable that prematurity would be associated with increased susceptibility to air pollutants including NO₂. Furthermore, a recent Institute of Medicine publication (Preterm Birth. Cause, Consequences, and Prevention) indicates the incidence of preterm birth continues to increase and represents a growing public health concern. Another susceptible group includes infants. A number of studies have demonstrated a relationship between exposure to NO₂ and increased incidence of SIDS (e.g. Ritz et al, Air pollution and infant death in southern California, 1989-2000, Pediatrics 118:493-502, 2006; Dales et al, Air pollution and sudden infant death syndrome, Pediatrics 113:628-631, 2004;

Klonoff-Cohen et al, Outdoor carbon monoxide, nitrogen dioxide, and sudden infant death syndrome, *Arch Dis Child* 90:750-753, 2005). The existing document cites Dales et al (Gaseous air pollutants and hospitalization for respiratory disease in the neonatal period, *Environ Health Perspect* 114:1751-1754, 2006) as an example of increased risk of respiratory disease among neonates exposed to air pollutants (though NO₂ was not associated). Genetic background as a susceptibility factor could/should also be better characterized. The few polymorphisms that have been evaluated for increased risk of susceptibility to NO₂ effects are only a beginning, and a more thorough examination of genetic contribution is needed. The current evidence for genetic component of host responsivity to O₃ is strong, and it is likely that genetic variants will also be important in response to NO₂.

Comments from Dr. Timothy Larson

Comments by T. Larson on Risk and Exposure Assessment to Support the Review of the NO₂ Primary National Ambient Air Quality Standard: First Draft and Draft Technical Support Document

General Comments: In general, these are well written documents that covers a lot of material in an efficient manner. I did not get the complete rationale for why all the analyses were necessary, but perhaps an overarching figure describing the process would be helpful at the beginning of the draft document. Given all the uncertainties in such an analysis, the approach used here is reasonable. If it turns out that on-road values in street canyons are systematically higher than those not in those canyons, the final exposure estimates may be low. A literature survey of this factor is recommended as a way to assess its importance.

Response to Specific Questions:

Air Quality Information and Analyses (Chapters 2, 5, and 6)

1. To what extent are the air quality characterizations and analyses technically sound, clearly communicated, appropriately characterized, and relevant to the review of the primary NO₂ NAAQS?

The air quality discussions in Chapters 2 and 5 are for the most part clear and to the point. I appreciated the relatively brief summary. The characterizations are based mainly upon the EPA data set which is a reasonable choice.

2. To what extent are the properties of ambient NO₂ appropriately characterized, including ambient levels, spatial and temporal patterns, and relationships between ambient NO₂ and human exposure?

The EPA data set is temporally rich and spatially poor. This point cannot be emphasized enough. We do not have many NO₂ monitors in most U.S. cities. Unlike NO₂ networks in many European countries, EPA has tried to site their monitors away from roadways in order to characterize the broader scale urban background values. This is not always successful. It is not clear if this factor has been accounted for in the data set.

Most of the studies that have reported simultaneous data from both near- and away-from-road monitors are from Europe. These analyses are potentially confounded by urban street canyon effects. Given that one of the main goals of the exposure exercise is to estimate near-road and on-road concentrations, I cannot tell if that factor has been considered in the choice of data.

One sentence that perhaps deserves more clarification is found on page 10, line 20. I am not sure what is meant by “the strength of the association varies considerably”. Do you mean the strength of the association varies considerably because of exposures from other microenvironments or because the experiments are not that precise and there is no association with any other factors?

3. We have evaluated air quality in a number of individual locations throughout the United States. What are the views of the panel regarding the appropriateness of these locations and on the approach used to select them?

The choice of locations is sensitive to the on-road estimates. Applying this model to New York is problematic, given the urban landscape that tend to trap the pollutants in street canyons. At least some of the monitors in Chicago have this same complexity. However, these cities were not chosen for further analyses, so I guess that choice seems OK.

4. In order to simulate just meeting the current standard, we have rolled up NO₂ air quality levels. To what extent is the approach taken technically sound, clearly communicated, and appropriately characterized? Do panel members have comments on the relevance of this simulation for reviewing the primary NO₂ NAAQS.

Linear roll-up of NO₂ rather than NO_x could be tricky, given that some of the NO₂ is directly emitted and some is formed immediately downwind. The downwind formation rate depends upon meteorology and upwind ozone, both of which are variable. There is also the complication that the recent adoption of catalytic converters on heavy duty vehicles results in more primary NO₂ relative to NO_x than in past years. This would imply that the NO₂ to NO_x ratios vary from day to day, by year, and with proximity to heavy duty vehicles. However, given all these uncertainties, there is really not much else to do. One could look at the Aermol line source predictions from Philadelphia, factoring in the variation in upwind ozone to see how much variability there might be in peak to annual mean ratios as a function of the annual mean.

5. Because of the impact of mobile sources on ambient NO₂, we have estimated on-road NO₂ concentrations. To what extent is the approach taken technically sound, clearly communicated, and appropriately characterized? Do panel members have comments on the relevance of this procedure for reviewing the primary NO₂ NAAQS?

On page 39, line 6, where it refers to a very strong near-road gradient that occurs within 10 meters of the roadway edge. Is this a typo? Did you mean 100 meters? It is well known that some although some NO₂ is directly emitted, some is formed immediately downwind. If the gradients are in fact that pronounced (i.e. 10 meters) near the road, small changes in the value at the EPA monitors that are located further from the road would make a big difference in the estimated on-road NO₂. In this case it would seem that the model is extrapolating outside the measurement space and therefore the sensitivity of the analysis results would depend strongly on the exact location of the EPA monitor relative to the road. I cannot tell because no sensitivity results are discussed.

6. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability?

The discussion of the uncertainty in the variables considered in the analysis is reasonable. I think that the near-road uncertainties are dominated by street canyon effects in built-up urban areas. This uncertainty could be qualitatively assessed at a minimum.

Exposure Analysis (Chapters 5 and 7)

1. To what extent is the assessment, interpretation, and presentation of the initial results technically sound, clearly communicated, appropriately characterized?

This is a rather difficult line of reasoning to follow. I would suggest a diagram showing how all the parts of the analysis fit together to achieve the desired goal. Otherwise, it is easy to get lost in details in one section that reads a lot like another one.

2. The draft risk and exposure assessment document evaluates exposures in Philadelphia. Future drafts will also evaluate exposures in Atlanta, Detroit, Los Angeles, and Phoenix. What are the views of the panel regarding the appropriateness of these locations and on the approach used to select them?

The first draft focuses on Philadelphia. This is a rather different city from many large cities in the U.S.; specifically it has a relatively high single family residential density. The distance from the census tract centroids to major roads is surprisingly large (median >400m). Is this typical? I think it may be on the high side, but I have no basis for comparison. I would think Los Angeles would be different and I know New York is quite different (90% of residents live within 100 m of a busy road). Given that we are only talking about a one hour exposure, and that the brief near-road exposures drive the high end of the hourly maximum ambient distribution, I would suggest making the model runs in cities based on the median value of the distance from the ct centroids to major roads.

3. Do panel members have comments on the appropriateness and/or relevance of the populations evaluated in the exposure assessment?

The asthmatic population seems like a reasonable choice, given the known health effects of NO₂. Is there information on the prevalence of asthmatics without managed care living near major roads (not necessarily because the roadway pollution created the asthma, but because of other demographic and economic factors).

4. To what extent are the approaches taken to model stationary sources and mobile sources technically sound and clearly communicated?

This section seems OK for Philadelphia. How do the Aermoc predictions of the relationship between monitor values and on-road values compare with the screening assessment values for m? Is the characteristic decay distance similar? Model comparisons with the annual average monitor values should also be presented as a scatterplot for all sites in the modeled cities.

5. Human exposures are modeled using APEX to simulate the movement of individuals through different microenvironments. Do panel members have comments on the microenvironments modeled?

Again, I would like to see some adjustment for street canyons. I think it is reasonable to assume that some individuals could spend a brief period of time in these microenvironments. There is some data on this in the literature.

6. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability?

The final results for the number of exceedances of the short term levels may be very sensitive to the near road enhancement factor that is derived from a model using census tract centroids. The results for Philadelphia would seem to underestimate these numbers, given the relatively large median distance of the population from roadways. In addition, given the non-linear decay of NO₂ concentrations away from roads, people living nearer the roadway within the census tract could be experiencing much larger short term exposures than others in the same tract. The opposite may be true but should not affect the final hourly NO₂ estimates as much if most of the exceedances occur in relatively small cts with relatively high population densities.

Comments from Dr. Kent Pinkerton

Characterization of Health Risks (Chapters 3 and 4, Sections 6.3, 7.8, and 7.9):

What are the views of the Panel on the overall characterization of the health evidence for NO₂? Is this presentation clear and appropriately balanced?

RESPONSE: The recognition of specific subpopulations that represent increased health risks to the effects of NO₂ exposure is a critical point. Pre-existing disease/symptoms (i.e., asthma) and/or infection risk, age and relative proximity to roadways are clearly important considerations. I think it is important to use all the existing data including epidemiology and human clinical studies with animal toxicology studies to provide biologic plausibility.

It is also important to consider what monitoring methods should be applied to determine what are the actual exposure conditions involved as health effects are observed. Further consideration must also include discussion of the appropriate exposure metric (i.e., 1 hour peak, vs. 24 hour average vs. annual average) to establish health effects due to NO₂.

Use of the animal toxicology literature to provide mechanistic insights into health effects is reasonable, based on the need to use higher than ambient concentrations of NO₂ to explain biologic plausibility to the observed human benchmarks of increased airway responsiveness, increased susceptibility to infection and exacerbation of asthma.

The characterization of health risks focuses on potential health benchmark values identified from the experimental NO₂ human exposure literature on airways responsiveness. What are the views of the Panel on using potential health benchmarks from this literature to characterize health risks?

RESPONSE: The determination of increased airway responsiveness with exposure to NO₂ at a level of 200 to 300 ppm for 30 minutes in asthmatics is clearly important. This data provides strong evidence of health effects for NO₂ at levels currently below the current NAAQS NO_x standard. Individuals at risk clearly include those with asthma for both children and adults who are asthmatic. Changes based on airway responsiveness, enhanced susceptibility to infection and/or asthma exacerbation are important considerations in advocating for health effects. Whether these considerations should arrive at the conclusion significant health effects with short-term exposure (0.5 hr peak levels) of NO₂ at levels within the range of 200 to 300 ppb or even lower levels (i.e., 40 to 80 ppb) due to children with asthma need to be tempered with assessment of all studies. However, based on the cited literature in the ISA draft document, benchmark levels of 200, 250 and 300 ppb for 1 hour seem highly appropriate.

In terms of cities selected for NO₂ monitoring, does Los Angeles present difficulty due to higher levels HNO₃ that may result in an overestimation of NO₂?

Do panel members have comments on the range of potential health effects benchmark values chosen to characterize risks associated with 1-hour NO₂ exposures?

RESPONSE: Again, studies which have suggested of increased airway responsiveness with exposure to NO₂ at a level of 200 to 300 ppm for 30 minutes in asthmatics are important. Epidemiological studies which have reviewed hospital admissions and emergency room visits have also provided growing evidence for NO₂ effects based on ambient average levels.

To what extent is the assessment, interpretation, and presentation of initial health risk results technically sound, clearly communicated, and appropriately characterized?

RESPONSE: In general, health risk results are well presented. It is critical that both epidemiology and human clinical studies both need to be considered in the overall assessment and interpretation of available information.

While the epidemiology literature will be considered in developing the Agency's policy assessment as part of an evidence-based evaluation of potential alternative standards, staff has judged that it is not appropriate to use the available NO₂ epidemiological studies as the basis for a quantitative risk assessment in this review. Do panel members have comments on this judgment and/or the rationale presented to support it?

RESPONSE: There appear to be ample scientific data of independent epidemiologic studies to show significant observed NO₂ effects that appear to remain robust when adjusting for multiple co-pollutants. It is understood that concerns of confounding effects remain, but should not be entirely dismissed for the purposes of a quantitative risk assessment in this review. The rationale presented to support this approach seems reasonable, based on the paucity of available cities/sites in the US meeting needed NO₂ levels for conducting risk assessment. However, it would seem critical to not minimize the epidemiological literature in consideration of risk assessment.

Comments from Dr. Armistead Russell

This document lays out the modeling approach EPA plans to use to calculate the number of individuals exposed to NO₂ levels of concern in association with varying potential standards. Specifically, the document lays out the areas that are to be modeled, approach to be used to calculate air quality, exposure modeling and estimated NO₂ exposures. They also assess some of the uncertainties and variabilities in the process.

My first comment is that, for the most part, it is an impressive effort. Seldom do they probably hear that maybe they are going too far in that I would hate to see the depth of their analyses limit the breadth, which ultimately might be of more interest. This should be taken as a compliment: I was impressed by the detail of the air quality modeling approach.

A first quibble is that the Introduction could be expanded to provide more of a picture of what was to come. A few paragraphs laying out the approach would be good, providing a flow of effort and information. Here they can define what models are to be used and why, as well as the specific outcomes of interest, and why. A second general comment is that the document is a bit uneven, with some sections being thorough and readily understood, while others lacked motivation and it was a bit difficult to see exactly what was done and why.

The first analysis is a so called “air quality data screen” used to characterize NO₂ at monitors in a number of areas in the US over 12 years, from which they choose a more limited set of metropolitan areas to be examined in some more detail to finally arrive at a workable number of locations to study in detail using air quality and exposure models. This section is thorough and achieves its objective of providing the data for choosing a set of locations to be studied in more detail. From the analysis, 15 locations were viewed as meeting the selection criteria, modified to 18 by additional issues. This number of locations is still not practical for complete exposure analysis.

The next section characterized the ambient monitors, with particular interest in their location relative to NO_x sources. NO_x sources of interest included roads and stationary sources. This section was thorough as well, though its need could have been better motivated.

Section 2.4 covered characterizing observed air quality in the selected areas from above, and characterized annual means, hourly concentrations and the variability in NO₂ levels in ten of the areas. A few quibbles with this section. First, more information on why the ten sites are chosen for presentation. Second, Figures 1-3 need work: units are needed on the vertical axes, and “spatial distribution” usually implies a map of concentrations. Table 7 should indicate what was being tested statistically. On page 14, they should actually say why they look at Philadelphia (since it is used later for more detailed study). A variety of statistical tests and plots are contained in this section, with a little description of what they are doing, but not overly motivated as to why. In many ways, when one got done with this section, one was left with the impression that much was done, but little was gained.

The initial approach to air quality simulation, as contained in 2.5, provides a set of simple procedures used to adjust air quality to adjust concentrations from historically observed levels to

just meeting the standard, as well as an approach to account for NO₂ levels near the road. While simple, the approach used to scale up levels to meet the standard is reasonable, recognizing that the PRB is minimal. This limitation should be noted. The method for estimating concentrations near/on roads is simple, but I was left wondering why? First, it is based on a rather slim set of data. Second, it is functionally wrong if one looks at standard Gaussian dispersion. The rate constant, k, is said to describe formation and decay, though it can not describe formation, and chemical decay of NO₂ over the length scale of interest is small. Through this whole section I was wondering why not use a dispersion model. In the end, the studies used to develop k were primarily from outside the US, and often for longer averaging times, the latter of which is particularly important. Table 12 should include the averaging times of the studies. They also need to spell out exactly how, in the final analysis, they will use the on-road factors calculated, and present it up front in 2.6 to motivate what is to come. As I was reading this, I was wondering if it would be used in the air quality modeling and exposure assessment, and was thinking, "I hope not." All told, while I am not thrilled with the method used, it is probably fine for how the final on-road factors are used as they are not central to the exposure modeling (I think...).

Section 2.7 on the estimation of benchmark exceedences was thorough, and it is here that one finally sees how the on-road factors are used. The final section was a listing of the likely uncertainties and processes leading to variability, which again was thorough, but not quantitative at all, and one is left wondering what is minimal, moderate and major in a more quantitative way. What does it really take to be major? Also, why does the uncertainty have to all be in the same direction to be moderate?

Some details:

P 52, 142: "...possible interferences." Further, it should recognize the 50% is extreme. Likewise, the vertical gradient ratios are extremes.

P 53, 18: ...monitors are sited...

Section 2.8.5: One could test the likelihood of overestimations by comparing the various years of data. I, too, suspect it is a minimal concern.

P 55, 120: Your approach assumes that a site <100 m from the road is impacted, so I would not be so tentative in the statement used.

P55, 142: do you mean accuracy instead of precision? Also, I am not sure how the bounds really get set. Please clarify.

Section 2.8.7: I would think this is the major uncertainty.

Section 3 gets to the exposure assessment. (Oh... a quibble, I would prefer approach versus methodology.) The introduction needs to give a short overview of the approach to motivate what is to come. The first task is using the prior analyses to pick a practical number of locations. Their criteria is sensible, and the final list is reasonable, though I would have chosen a high

elevation city in a Rocky Mountain state (e.g., Denver, Provo) instead of Phoenix, given the proximity between and similarities in Phoenix and Los Angeles.

As noted above, I thought that the analysis in this section, as applied to Philadelphia, was a tour de force. The model choices (AERMOD, APEX) are appropriate, and they have gone through a very extensive data development procedure. I might argue that they should not calibrate the AERMOD results, as I would prefer an evaluation of the results and let that guide further consideration. They need to explain with mathematical equations, how they calculate the “local concentrations” (Page 89), and then how those are used. My major concern with the application of APEX is that there is no real evaluation. I think the state of this section bodes well for things to come.

Minor comments on section 3: Are you sure commercial air craft do not contribute more NO_x. In Atlanta, our estimated NO_x emissions from aircraft (4910 tpy) are about 10 times the GSE.

P133: 11: Use practical, not possible.

Comments from Dr. Jonathan Samet

General Comments:

This first draft of *Risk and Exposure Assessment* attempts to link the findings of the ISA on risks to health and population patterns of exposure to human health risks under various scenarios of ambient NO₂ concentration. The document is still “in progress” with a still incomplete exposure characterization. In developing the document, the Agency faced the challenge of linking the annual standard to temporal profiles of exposure that are far briefer, i.e., one hour and relevant to the selected health outcome measure. The result is an extensive series of assumptions and models. The document is difficult to follow as a result and presentation needs to be improved. At the minimum, I would propose that an introductory section be developed that lists out in tabular or graphic form the approach taken, both with regard to the chapter entitled “Ambient Air Quality and Health Risk Characterization” and the subsequent chapter “Exposure Assessment and Health Risk Characterization”. The reader is challenged to follow the multiple steps and assumptions in these analyses.

In selecting the concentrations of concern, the Agency bases its choice on the observations with regard to airways responsiveness, while noting other short-term effects. There needs to be a careful consideration of the clinical and public health significance of the effects observed in the short-term studies that have identified the association of ambient NO₂ with increased airways responsiveness. Chapter 4 of the draft ISA sets out general considerations, but the ISA does not deal specifically with the significance to individuals or to populations of a transient increase in airways responsiveness. Guidance is needed as a basis for using the risk and exposure assessment for policy purposes.

At this point in the evolution of the document, explicit consideration needs to be given to whether the exposure characterization is sufficiently certain to be useful and whether the approach should be completed for the other designated cities. The limitation of the AERMOD output is apparent, with substantial adjustments needed when model outputs were compared to actual monitoring data. Additionally, the exposure characterization using APEX is subject to numerous uncertainties. The results from Philadelphia are informative on the potential for exposures in ranges that may affect the key health indicator. Will completing this characterization for other locations add substantially to the information base needed for decision making? I note that Section 7.10 lists numerous sources of variability and uncertainty but reaches no “bottom line” on the overall level of uncertainty. This summary judgment is needed to inform utilization of the results.

Characterization of Health Risks:

My comments above address many of the principal issues around the characterization of the risks to health. I concur with the decision not to use risk estimates from the epidemiological studies. While there is mixed evidence on the effects of NO₂ on airways responsiveness, an increase is plausible and documented in some studies. The percentage of persons with asthma who are potentially susceptible is not known, an uncertainty that

should be acknowledged. In fact, in the exposure characterization and health risk analysis before Philadelphia, all persons with asthma are assumed to be susceptible to NO₂, which may not be the case (I also note that the percentages of adults and children assumed to have asthma appear to be somewhat high and no source is given for the percentages used).

I have no specific comments with regard to Sections 6.3, 7.8, and 7.9, beyond those made above.

Specific Comments:

Page #	Line #	Comment
10	3-6	In regard to what standard and for what purpose?
10	10	“introduce uncertainty...” Of what sort?
10	13	This would be expected
13	1	“highly susceptible..” What does this mean?
28	11	Sentence not clear.
28	14	No- use of air quality concentrations as surrogates
33	23-25	Outdoor <u>vs</u> personal
90	20	“summary of the (utility) of the estimated...”
92	11	Based on?

Comments from Dr. Richard Schlesinger

Section 3.1. There appears to be some confusion over use of the terms “susceptible” and “vulnerable.” Both terms are used for specific populations, namely children and the elderly, when it is indicated that there is age related susceptibility as well as vulnerability. Based upon the definitions of the two terms given in this section, children and the elderly should be considered as susceptible populations rather than vulnerable populations.

p. 15, line 16-17. Perhaps the sentence should read “...NO₂ may increase an allergen-induced increased airway responsiveness...” rather than “inflammatory response.”

p. 20, line 5. There needs to be a clearer justification for use of the lowest benchmark level of 0.2 ppm inasmuch as this is below the lowest level used in controlled human studies at which effects were seen.

Comments from Dr. Christian Seigneur

The exposure and risk assessment methodology that was reviewed earlier appears to have been properly implemented and the First Draft presents a clear and detailed description of the results to date. The detailed exposure modeling has only been reported for one metropolitan area, Philadelphia, so far but the results for that area provide sufficient information to evaluate the implementation of the methodology.

My main concern with the results presented for Philadelphia is the poor performance obtained when comparing the air quality modeling results with ambient NO₂ concentrations. This comparison is presented in Table 26 on p. 91. Performance appears to be satisfactory for two monitors (292 and 471) since the model simulation results are within 4 to 35% of the measurements. However, performance is extremely poor at the third receptor (043) with underestimations on the order of a factor of 3 to 4. Also the year-to-year variability is not predicted correctly at one of the receptors (471) where the measurements show a 17% decrease from 2001 to 2003 and the model predicts a 30% increase. Clearly, the significant underprediction at receptor 043 is the major concern. EPA does not explicitly address this receptor-specific underestimation but instead treats it as a regional underestimation, which is inappropriate because underestimations are significantly less at the other two receptors. The significant underestimation at receptor 043 suggests that a local source (or sources) has not been taken into account in the emission inventory (or has been significantly underestimated). This source affects primarily receptor 043 and does not affect significantly the other receptors (since underestimations are much less at those other receptors). Therefore, adding a “background” concentration that is uniform across the area does not correct the problem (column titled AERMOD Final): NO₂ concentrations are then slightly overestimated at receptors 292 and 471 and they are still significantly underestimated at receptor 043 by factors of 1.3 to 1.9. Furthermore, the spatial distribution of predicted NO₂ concentrations is still incorrect with concentrations at receptor 043 that are 1.7 to 2.5 times smaller than at the other two receptors, whereas the measurements only show differences of a factor of 1.2 or less. Such poor model performance results cast doubt on the robustness of the subsequent analysis since exposure in the vicinity of that receptor could be off by a factor of two. I recommend that EPA carefully diagnoses the causes for the model underprediction at receptor 043 and either correct the model inputs such as the emission inventory (the preferred approach) or make a post-modeling correction that accounts for this receptor-specific discrepancy.

Another point related to the AERMOD performance evaluation pertains to the measurements used to evaluate model performance. Table 26 lists only three receptors but Table 25 lists 10 monitors where NO_x measurements are available. Model performance should be conducted with all the measurements available.

The details of the application of AERMOD to NO_x emissions need to be presented. Was AERMOD simply applied to NO₂ emissions or was AERMOD applied to both NO₂ and NO emissions with some oxidant correction to account for the conversion of NO to NO₂?

Another aspect of the analysis that needs to be revised is the use of benchmark scaling. I understand that this approach is computationally more efficient than the alternative of

redoing the calculations with scaled-up NO₂ concentrations. If this approach seems valid when performing the exposure analysis with air monitoring data, it seems inappropriate when applied to a model that combines contributions from outdoor air (i.e., the component being scaled) and indoor air (i.e., the component that is not scaled). It seems that for the exposure model that combines both outdoor and indoor air exposure, the calculations need to be redone with only the outdoor concentrations being scaled up to the current NO₂ standard.

The discussion of uncertainty and variability needs to be improved to provide the reader with some semi-quantitative information on which sources of uncertainty/variability are the most likely to be significant and to what extent they could affect the results of the assessment (i.e., within 10%, a factor of 2, an order of magnitude?). For example, the discussion of the interference problem for NO₂ measurements (section 6.4.2, p. 60) could point out that the largest errors occur during summer and at locations far downwind of sources, i.e., in most cases in locations which do not have the highest NO₂ concentrations. Also, the discussion of uncertainties in Section 7.10 only addresses data uncertainties. The uncertainties associated with the formulation of the models (atmospheric dispersion model, microenvironment model) also need to be discussed. References to AERMOD model performance (Perry et al., “AERMOD: A dispersion model for industrial source applications – Part 2”, *J. Appl. Meteorol.*, 44, 694-708, 2005) and the performance of roadway dispersion models (Benson, “A review of the development and application of the CALINE3 and CALINE4 models”, *Atmos. Environ.*, 26B, 379-390, 1992) would be helpful.

The assumption was made that the NO₂/NO_x emission ratio was 10%. Power plants typically have a much lower ratio and some mobile sources may have a much larger ratio (see discussion of retrofitted diesel engines in the ISA). It would be useful to conduct a sensitivity analysis where this ratio is modified to assess its impact on the results of the analysis.

Editorial comments:

On p. 24, line 22 and p. 25, line 26: The relationship should be causal rather than casual.

On p. 38, line 22 (and in the TSD): Rate constant generally refers to a change with time; I suggest “decay constant” or simply “constant”.

On p. 70, line 25: Delete the first “area”.

Comments from Dr. Elizabeth “Lianne” Sheppard

Summary:

The current exposure assessment is problematic because it is most likely underestimating the exceedances. It is possible (but questionable whether it can succeed) that the improvements to the work that has already been done will suffice to provide good estimates of exceedances. In comparison to O₃, NO₂ is more difficult to model for exceedances. Monitored NO₂ is more spatially variable than monitored O₃ (due to monitor siting criteria). It is extremely difficult to predict NO₂ at the one-hour average time scale (preferred is the 8-hour time scale for O₃, better yet would be the 24-hour average time scale). Thus I believe that the evidence from the two exceedance calculation exercises (monitored data and APEX) could be fatally flawed and therefore basing quantitative assessment on exposure assessment will bias conclusions towards no need for a short-term NO₂ standard. I recommend taking one of two approaches: 1) dropping quantitative exposure assessment for NO₂ until better spatio-temporal prediction methods are available for the 1-hour time scale and there is better justification that the predicted exposures capture the upper tail of the concentration distribution or 2) redo the exceedance evaluation to only consider the 24-hour average time scale.

I have more confidence in the risk assessment for NO₂ based on epidemiology studies (in contrast to the exposure assessment). I would prefer the asthma panel study data be used for risk assessment, but I understand it may be difficult to obtain baseline risk information. The time series study results may provide the best basis for the quantitative risk assessment due to a combination of practical and scientific considerations. These studies are based on 24-hour average “usual” population exposure to ambient concentration and will be less sensitive than the exposure assessment due to the long (24-hour) averaging time and the ability to focus on monitors that are more representative of the majority of the population (i.e. not those near roads). The challenge may be to obtain baseline rate information for populations of interest. An alternative may be less well-founded estimates coupled with a sensitivity or ensemble analysis that considers a range of different estimates. I think the limitations of NO₂ time series studies are not significantly worse than those for O₃ and PM and so such limitations should not be cited as reasons for not conducting a quantitative risk assessment. I think relative risk estimates can be applied to a given area even if they were not obtained locally or from a multi-city study. Overall I think some a quantitative assessment (based on epidemiological studies) should be done.

Air quality information and analysis (ch 2, 5, 6):

Because the exceedance evaluation is focused on the extremes of the distribution, choices must be made to get the distribution upper tail correct and assessments done to show the extremes have been captured well. NO₂ is spatially highly variable, particularly on the 1-hour average time scale. Typical data analyses focus on the mean and don't worry about the extremes. We don't have that luxury in the context of estimating exceedances. This is a *crucial point that pervades all of the ambient data analysis and modeling and the entire approach to exposure assessment*. This concern *must* be addressed.

- Basing exceedance estimates on monitor-years of data is problematic. Monitors are not sited to represent to population ambient-source exposure within a location. The air quality-based exceedance evaluation could be seriously undercounting

exceedances because monitors are not sited in the highest concentration areas in proportion to the population living near such sources. Furthermore, within an area monitor-years are not exchangeable, but the analysis approach appears to treat them as exchangeable. It matters whether a near-road site is included or excluded in a given year.

Exposure analysis (Ch 5, 7):

The APEX analysis suffers from the same potential problem with lack of variability of the predicted ambient concentration data. *All evaluations must assess whether the tails of the predicted distribution are correct.* Simple mean adjustments are woefully inadequate.

- The APEX modeling approach is very thorough and addresses many (but not all) important sources of variability. The addition of the algorithm to incorporate day-to-day correlation of activities within individuals is a good enhancement, and an important feature given one of the summaries is the number of repeat exceedances within a person.
- Verify the AERMOD predictions are aligned with data with respect to their variability, not just their mean (see e.g. Table 26 p. 91). If there is inadequate variability of the predictions, the number of exceedances will be underestimated.
 - By not including all the NO₂ sources in an area, the net effect of the AERMOD approach should be to dampen the variability. This cannot be corrected by a fixed increase in the mean. It is important to appropriately model variability if we are going to correctly capture variability.
 - An additive mean adjustment won't affect the variability.
 - Explain how the correction at monitoring sites affects predictions at other locations without monitors.
 - At the 3 monitors, present bi-variate analyses such as scatterplots of paired hourly predicted vs. measured concentrations. Other figures, such as time series plots of the data and differences between predictions and measurements will be useful as well.
 - Features of the spatio-temporal NO₂ distribution are not being captured by the current approach. For instance, high exposures should be defined as a function of building geometry and urban centers, not census centroids.
- The diurnal cooking pattern section description could be clearer (p 102). I'm concerned that the approach may smooth NO₂ exposure too much.
- In Section 7.9 I think the presentation would be much clearer if the assumption is stated up front that exposure to asthmatics and non-asthmatics is the same.
- I suggest focusing on the adequacy of this approach in Philadelphia rather than moving to other cities. It would be better to document what the approach is missing and conclude that it should be discounted/discontinued rather than moving forward with replication of a potentially misleading approach.

Health risks: (Chapters 3 and 4 and Sections 6.3, 7.8, and 7.9):

I recommend using the epidemiological study results to do a true quantitative risk assessment. Since this analysis will be based on time series studies with 24-hour average concentration data, many of the issues mentioned above of properly modeling the tails of the exposure distribution will not be a problem.

Specific response to charge question 2. *The characterization of health risks focuses on potential health benchmark values identified from the experimental NO2 human exposure literature on airways responsiveness. What are the views of the Panel on using potential health benchmarks from this literature to characterize health risks?*

This characterization is limited by the feasibility of controlled human studies. The sickest and most susceptible individuals can't be studied in this setting, and highly responsive but perhaps not that clinically meaningful health endpoints are selected. Such endpoints are likely to show a response during a short-term exposure but they may not be too meaningful from a clinical perspective.

Overall reporting:

- In looking ahead to the use of these data, it is important to emphasize that the estimates come from a small subset of the US and thus would have to be scaled up in order to reflect the entire US population. I am concerned that the quantified estimates may be used “as is” without appropriately reflecting the geographic areas and populations they actually represent. In scaling up, consideration needs to be given to the fact that locations were selected because the data suggest they have high exposure. Some of this may be due to the data that are available, but some may also be due to the areas not selected having fewer high levels of NO2.
- I iterate once again that statistical significance does not equal scientific importance. Table 1 implies that it does and needs to be formulated to remove that focus (e.g. replace with effect estimate and CI). Remove this feature (of focus on statistical significance as the key feature to summarize) whenever it appears throughout the document.

Additional specific comments: ERA

- p 27 Add summary and conclusions
- p 30 l 21-24: This statement is too strong. Monitor siting limits the types of locations measured.
- p 31 l 21: Why not assess whether we have a NO2 health problem under current conditions?
- p 31 l 11-12: It is the COV that is shown to be relatively constant, not the variance.
- p 31 l 21: Make it clear that the maximum is of all annual averages from monitors reporting data for a particular city and year. Include the number of monitors in the supporting summary table (Table 10 of the TSD).
- p 39 l 7: You mean 100 meters, not 10?
- p 54 l 2-4: This is the key problem with doing this monitor-based analysis. It matters a great deal which site-years are included. It is impossible to generalize without knowing a whole lot more about the data and the representativeness of the monitor siting relative to population exposure.
- p 54 l 9-10: Likely this is an unreasonable assumption.
- p 61 l 8: I would judge bias and uncertainty could both be huge.
- p 61 l 11: Ignoring monitoring objectives and land use means the analysis is weighted in favor of the most popular siting criteria.
- p 65 Table 6: Several uncertainties are severely downplayed, particularly spatial representation and scale, model choices

- p 91 Table 26: Completely inadequate evaluation. Look at variation as well since the focus is on exceedances. Assess the data on the time scale of interest.
- p 91 l 11: How do asthmatics differ with respect to exposure? p 107 l 17: Why not just assume exposure is the same for both asthmatic and non-asthmatic individuals? If there is good reason to keep these distinct, note how it is different.
- p 97 l 17: Table 28?
- p 100 Table 30: What does N refer to?
- p 102 discussion of diurnal cooking: This isn't very clear to me; I am concerned it may smooth out NO₂ exposures too much.
- p 112 l 15: Of how many? Unqualified numbers like this taken out of context are likely to be misused.
- e.g. p 120 Figure 14, 15: I would prefer tables.
- p 126 Section 7.10.4: The certainty of the air quality modeling is highly overstated.
- p 131 l 21-22: The lack of this information suggests systematic underestimation of exceedances.

TSD:

- Shift the focus of this document:
 - Use modern software to make convincing pictures of raw data (e.g time series plots with smooth curves overlaid)
 - De-emphasize statistical testing unless it changes your actions
- Sections 2.1-2.4: Do analyses to convince me your approach is adequate. So far I am not convinced.
- Section 2.2: Define location in two ways (geographic space and design space (i.e. covariates)). Revise the analysis accordingly. Focus on exploratory analyses that help determine which aspect is most influential.
- p 6 Table 3: Add number of valid monitors and monitor-years. Stratify by siting/design characteristics (e.g. near vs far from road).
- Section 2.4:
 - Revise so there is less emphasis on testing, more emphasis on data description.
 - I want to see maps, time series, exploration of design/siting features (help discover which are most important).
 - Focus on the questions of interest. Here are mine:
 - What time of day are exceedances most likely to occur?
 - What locations are exceedances most likely to occur
 - How similar are cities, locations within cities?
 - What is the pattern of the current standard over space (region) and time, siting/design characteristics?
- p 10 starting l 41: What is the purpose of all this statistical testing? Are the assumptions justified? (It is not good enough to just state the assumptions.)
- p 11 l 31-34: And the fraction of sites influenced by roads and point sources.
- p 12: How do these figures show spatial distributions? Does the ordering of the x-axis reference space? How about replacing these with maps?
- p 15 l 1-3: This means you can't simplify, i.e. you can't treat the locations as exchangeable.

- p 15 Figure 5. The x axis has no meaning. Represent space by showing geography and/or design.
- p 15 Table 8: Make the table heading clear.
- Section 2.4.4: Clarify. lines 3-4: Not really. Lines 5-6: Does this adjust for missing monitors? It appears the order of the curves is fairly consistent for the top half of the distribution.
- p 19 1 5 “confirmed”: Huh? The comments on the figures and the statistical testing results conflict. 19: If the assumptions of the test aren’t met, why is it being used?

TSD Appendices:

- Appendix B: Define N in table headings. Do an analysis that shows individual time series (plot data, overlay smooth) and makes comparisons by siting characteristics

Comments from Dr. Frank Speizer

Risk and Exposure Assessment to support review of NO₂ Primary NAAQS: First Draft, April 2008

Page 20-21, Para beginning on line 22. There are several risks reported in this paragraph that are substantially different from each other, given that they are reported for a 30 ppb and 20 ppb 1 hour exposures. Simply reporting them seems not enough. Some explanation about the differences should be discussed.

Page 22-24: Short term effects: With all due respect it seems to me with over 50 peer reviewed studies since the last assessment and the consistent mechanisms shown in different toxicological studies that staff is wrong to conclude that a quantitative risk assessment is not warranted. I do not understand how staff can be so sure that the judgment would not meaningfully inform the administrator, particularly when there will be a new administrator, who hopefully will be more independent of the OMB in reading the science. At the very least a proper risk assessment even if not conclusive will hopefully point us in the direction for future work, so that we do not have to wait for another 50 studies in the next 5 years that will take us no further than we are now!

Page 33, section 5.4.1, end of para. The logic of this is not readily clear to me. For the Boston example the highest reading is 30.5 ppb. What is logic of applying the F factor to scale up to .053 ppm? After all the .053 ppm is the max administratively derived value and thus arbitrary. Should this be an arbitrary value for everywhere? If there are real measurements, why not use them?

Page 33, Para 5.4.2 Not clear to this reader why roll up and roll down is used. Is it simply to help with the computer simulation or is there specific logic that makes the procedure meaningful? It becomes more confusing when the average concentrations of 33 ppb become because of the F factor 126,157,189 ppb.

Chapt. 6, page 34 sentence ending line 10. An alternative would be to consider introducing the need for a short term standard. This should be discussed.

Table 2, Page 37: What does it mean to have sites with less than 15 complete measurements/year over a 5 year period? Doesn't completeness depend upon how many sites in each city?

Comments from Dr. George Thurston

In these pre-meeting comments, I will focus upon responding to my assigned questions for the REA.

Characterization of Health Risks (Chapters 3 and 4 and Sections 6.3, 7.8, and 7.9):

1. *What are the views of the Panel on the overall characterization of the health evidence for NO₂? Is this presentation clear and appropriately balanced?*

This document presents a reasonable concise summary of the evidence presented in the NO_x ISA.

2. *The characterization of health risks focuses on potential health benchmark values identified from the experimental NO₂ human exposure literature on airways responsiveness. What are the views of the Panel on using potential health benchmarks from this literature to characterize health risks?*

This benchmark analysis is fine, as far as it goes. However, see my remarks below about the levels considered, and the need to also consider NO_x epidemiology.

3. *Do panel members have comments on the range of potential health effects benchmark values chosen to characterize risks associated with 1-hour NO₂ exposures?*

Considering effects only as low as 200 ppb seems incomplete, given that the 2nd Draft of the NO_x ISA concludes (on page 5-22) that “In studies that have examined the concentration-response relationships between NO₂ and health outcomes specifically, there is little evidence of an effect threshold.” Lower benchmarks, closer to ambient, are needed, if population effects are to be more realistically modeled.

4. *To what extent is the assessment, interpretation, and presentation of initial health risk results technically sound, clearly communicated, and appropriately characterized?*

I don't have any problem with the presentation, just the assessment itself, which is far too limited in scope.

5. *While the epidemiology literature will be considered in developing the Agency's policy assessment as part of an evidence-based evaluation of potential alternative standards, staff has judged that it is not appropriate to use the available NO₂ epidemiological studies as the basis for a quantitative risk assessment in this review. Do panel members have comments on this judgment and/or the rationale presented to support it?*

Yes, I have a major problem with this approach. It is incomplete. The EPA staff notes (on page 23) that:

“The preferred approach for conducting a risk assessment based on concentration-response relationships from the epidemiological literature would be to rely on studies of ambient NO₂ conducted in multiple locations throughout the United States that employ both single-pollutant and multi-pollutant models. This approach would provide a range of concentration-response functions that are relevant to specific cities in the United States.”

Moreover, in the NO_x ISA (on page 5-8) EPA states that:

“Taken together, recent studies provide scientific evidence that NO₂ is associated with a range of respiratory effects and are sufficient to infer a likely causal relationship between short-term NO₂ exposure and adverse effects on the respiratory system. This finding is supported by a large body of new epidemiologic evidence, in combination with findings from human and animal experimental studies”.

But the REA now completely ignores this conclusion of the ISA, dismissing the epidemiological evidence as too weak for application in the REA. These two documents are seriously conflicted, and do not now make sense together. This should now be rectified by recognizing the need for the application of the epidemiology results to the REA. The fact that many of the analyses do not have multi-pollutant models is not a barrier, as multi-pollutant models are only useful as sensitivity analyses, not for the development of dose-response estimates. This is because the regression betas of multi-pollutant models are not Best Linear Unbiased Estimates (BLUE), given the oftentimes high inter-correlations present between the estimates. Therefore, the appropriately conservative public health estimates to use for the risk assessments are the single pollutant coefficients, anyway. So this is not the barrier that the EPA asserts that it is. An application of the epidemiological evidence for respiratory effects of NO₂ to the risk assessment at ambient levels is absolutely required, or the EPA will not have met the objectives of this document.

Comments from Dr. James Ultman

The explanation of the methodology lacked clarity, particularly in the estimation of on-road NO₂ concentrations and the alternative roll-up method used in the exposure analysis.

In section 5.4.2, the explanation of the adjustment of the health effect benchmark level is not complete. It is important to state the conditions under which theoretical conditions this adjustment is appropriate. In particular, the personal exposure models have to be linear with respect to ambient concentration and indoor exposure sources must be comparably adjusted. Non-linearity in the model could arise, for example, if activity patterns of the subjects depended on ambient air level(at high ambient concentrations, asthmatics may tend to exercise less outdoors and spend more time in indoor environments).

It appears to me that a simple exponential extrapolation was used to estimate on-road conditions from off-road monitor measurements. Yet, equations 2 and 3 are confusing to this reader. For example, it is not clear why is C_b defined as “NO₂ concentration at a distance from the roadway, not directly influenced by road or emissions.” The source and assumption in equations 2 and 3 needs to be better explained.

Along the same lines, it would be useful to include a graph showing the distribution of the decay factors, k and a table summarizing the data that served as a source of these values (see also my comments on chapter 2 of the ISA). I am not convinced that this distribution constructed from a variety of different locations is directly applicable to Philadelphia.

The ambient air predictions of the AIRMOD model need to be better calibrated. The current approach is to use a constant baseline correction to force the annual-averaged value of the simulated receptor concentrations averaged over all receptor sites to agree with the corresponding measured values derived at the corresponding monitoring sites. Assuming linearity in the calibration, perhaps it would be possible to specify both slope and intercept calibration constants for separate monitors by comparing simulated and measured hourly-average concentrations.

Comments from Dr. Ronald Wyzga

Overall Comments:

The authors of this document are to be congratulated for producing a clearly written document on a complicated subject. The analyses undertaken are very complex and obviously are the result of considerable forethought and work. Given this complexity assumptions have been made that are for the most part reasonable. Some of my colleagues have raised issues about some of these assumptions, and these clearly merit further discussion. Where there is uncertainty in the assumptions I would like to see the effects of uncertainty embedded in the results. I believe range estimates are appropriate and can convey reality more clearly than a point estimate accompanied by explanatory text.

Charge Questions:

(Chapters 2, 5, and 6)

1. To what extent are the air quality characterizations and analyses technically sound, clearly communicated, appropriately characterized, and relevant to the review of the primary NO₂ NAAQS?

The methods were clearly communicated and clearly relevant.

I am uncomfortable with the approach when modeling the scenario for just meeting the standard. That approach assumes that all monitoring stations just meet the standard, a scenario that the document acknowledges to be highly unlikely. I believe that a more forthright approach is to acknowledge the uncertainty associated with the scenario and to present a range estimate for which the above scenario (every station just meeting the standard) provides an upper bound and the lower bound would be estimated by applying the “as is” scenario to all stations meeting the standard and reducing the ambient concentrations to the standards where they are exceeded. This range would be large, but it is more realistic than the estimate presented in the report. That estimate is misleading because it is most unlikely that NO₂ concentrations would increase to the standard level in areas where they are currently below the standard. It is conceivable they could increase in some areas, but with envisioned NO_x controls, universal increases are most unlikely.

2. To what extent are the properties of ambient NO₂ appropriately characterized, including ambient levels, spatial and temporal patterns, and relationships between ambient NO₂ and human exposure?

I think the relationships are clearly and reasonably summarized. More attention could possibly be given to the uncertainties associated with some of the relationships.

3. We have evaluated air quality in a number of individual locations throughout the US. What are the views of the panel regarding the appropriateness of these locations and on the approach used to select them?

The approach is reasonable.

4. In order to simulate just meeting the current standard, we have rolled up NO₂ air quality levels. To what extent is the approach taken technically sound, clearly communicated, and appropriately characterized? Do Panel members have comments on the relevance of this simulation for reviewing the primary NO₂ NAAQS?

I have problems with this approach. See my comments on the first question above.

5. Because of the impact of mobile sources on ambient NO₂, we have estimated on-road NO₂ concentrations. To what extent is the approach technically sound, clearly communicated, and appropriately characterized? Do Panel members have any comments on the relevance of this procedure for reviewing the primary NO₂ NAAQS?

This approach is reasonable although uncertainties could be addressed more explicitly.

6. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability?

I like Table 16; at issue is whether some of the uncertainties should be examined in more detail to present range estimates. In particular the results of sensitivity analyses could be presented in those cases where the magnitude of the bias has the potential to be moderate.

(Chapters 5 and 7)

1. To what extent is the assessment, interpretation, and presentation of the initial results of the exposure analysis technically sound, clearly communicated, and appropriately characterized?

These chapters are well-articulated; the authors are to be complimented for the clarity of presentation for such a complex analysis. It is clear that a lot of thought and work went into the presented analysis. My only concern is that the uncertainties are not embedded into the estimates of exceedances presented. There is uncertainty associated with the various estimates and these should be presented along with point estimates. Also see my comments above on the treatment of the treatment of “just meeting the current standard” scenario. I would very much prefer a range estimate for this case.

2. The draft risk and exposure assessment document evaluates exposures in Philadelphia. Future drafts will also evaluate exposures in Atlanta, Detroit, Los Angeles, and Phoenix. What are the views of the Panel regarding the appropriateness of these locations and on the approach used to select them?

The selection of these appears to be appropriate. If resources become a concern in subsequent analyses, I would be comfortable with the consideration of Philadelphia, Detroit and Los Angeles. Analyses of these 3 cities would portray the extent of risks associated with alternative standard levels. They appear to be the worst-case scenarios.

3. Do Panel members have comments on the appropriateness and/or relevance of the populations evaluated in the exposure assessment?

These appear to be appropriate and the most important populations to consider.

4. To what extent are the approaches taken to model stationary sources and mobile sources technically sound and clearly communicated?

My concern is that the uncertainties associated with models are not always clearly articulated.

5. Human exposures are modeled using APEX to simulate the movement of individuals through different microenvironments. Do Panel members have comments on the microenvironments modeled?

They are all reasonable; experience may suggest that some of them could be eliminated in other cities.

6. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability?

See comments on the first question of this section.

(Chapters 3,4 and section 6.3,7.8, and 7.9)

1. What are the views of the Panel on the overall characterization of the health evidence for NO₂? Is this presentation clear and appropriately balanced?

This is reasonable.

2. The characterization of health risks focuses on potential health benchmark values identified from the experimental NO₂ human exposure literature on airways responsiveness. What are the views of the Panel on using potential benchmarks from this literature to characterize health risks?

The use of benchmarks is appropriate although it would also be useful to find some way to include the probability of response associated with a benchmark; i.e., all subjects do not respond at levels just above the benchmark.

3. Do Panel members have comments on the range of potential health effects benchmark values chosen to characterize risks associated with 1-hour NO₂ exposures?

I believe the range is appropriate; one of the difficulties is that some benchmarks are associated with more adverse endpoints than others.

4. To what extent is the assessment, interpretation, and presentation of initial health risk results technically sound, clearly communicated, and appropriately characterized?

Again the document is well-written and clear. I would like to see greater use of range estimates to characterize some of the uncertainties in the analyses.

5. While the epidemiology literature will be considered in developing the Agency's policy assessment as part of an evidence-based evaluation of potential alternative standards, staff have judged that it is not appropriate to use the available NO₂ epidemiological studies as the basis for a quantitative risk assessment in this review. Do Panel members have comments on this judgment and/or on the rationale presented to support it?

Given the complexities associated with the results from epidemiological studies, I believe it is reasonable to use the results from clinical studies. Given the wide variety of results and specific model details across epidemiological studies, it would be difficult to decide which studies to use as a basis for dose-response. For that reason the current approach is more defensible

Specific Comments:

p. 34, l. 5: "These" air quality data

p. 97, l. 17: Table 28