Dr. Lianne Sheppard

For my preliminary comments I am only focusing on major points I have identified in advance of the meeting. I will include additional points in my final comments, including more careful review of parts of Chapter 2.

Organization and clarity

Overall the organization of the document is very good and much better than the 2008 NOx documents. Some key elements that I have appreciated are:

- Inclusion of the Preamble to clearly put the objectives of the ISA and the review process into context.
- Division of summaries into the executive summary, longer chapter 1 and results-specific summaries is helpful (though a bit repetitive for anyone reading multiple summaries in one sitting – I think this is unavoidable and the inclusion of multiple types and levels of summary is needed). With both the Executive Summary and overview Chapter 1 readers get a good overall perspective of the evidence and conclusions.
- Table 1-1 is a good overview of results for inference
- Integration of evidence from animal and human studies as a function of endpoint.
- Good discussions of the evidence in the context of the causal conclusions that are drawn.
- Well-designed tables that focus on the information needed for causal conclusions.
- Great cross-referencing of the document facilitating navigation.
- Excellent and easily accessible supporting information by integrating the HERO database

Exposure modeling and exposure measurement error

One of my major suggestions is that better/different attention be paid to exposure modeling and the concept of exposure measurement error, particularly in the context of epidemiological studies of long-term exposures where the focus is on spatial exposure variation. I believe that scientific understanding of the role of exposure in epidemiological inference to be at the cusp of reaching a deeper level of insight and I suggest that recognition of the potential of the emerging insights be incorporated into this document. I think such a discussion is even more important for NOx than for PM because NOx is a much more spatially heterogeneous pollutant and thus has more potential for epidemiological study findings to be impacted by the details of the exposure modeling. In my revised comments I will incorporate an updated perspective based on careful review of the relevant sections in Chapter 2. Here are some specific suggestions based on my reading of Chapter 5:

1. I suggest incorporating better summarization of the exposures used in the long-term epidemiological studies into the document. Results tables should incorporate more than just the type of exposure model used.

2. There should be some perspective included on the epidemiological inferences that can be drawn from the diverse set of exposure modeling strategies used in the cited papers (from e.g. nearest monitor, land use regression, dispersion modeling). There aren't yet any definitive statements that can be pulled from the existing literature, but I think the discussion can be broadened to reflect the dynamics of the exposures used in many studies and the aspects of them that may affect inference. Here are some: type of exposure model (most notably contrasting those that rely on measurements vs. physical and/or chemical models alone), spatial extent of the study and monitoring network, source of the monitoring data (e.g. regulatory network only or study-specific measurements), simplifying assumptions inherent in the work (e.g. are 2-4 weeks of data assumed to represent an annual average?), approach to smoothing/modeling over space (focusing on whether the model is "up" to capturing the sources of spatial heterogeneity in the pollutant), alignment of the monitoring and subject locations, size of the monitoring network (i.e. number and density of monitors used to develop the exposure model) and monitor siting criteria (e.g. are specific locations systematically omitted due to regulations?).

3. There should be some direct statements about the importance of the relatively high spatial variability of NOx in the evaluation of exposure assessment for epidemiological study inference. Unlike PM, which is spatially a much more homogeneous pollutant, the approach to exposure modeling of NOx and the set of monitors used in a given study, with respect to their numbers and locations, could have a major impact on the inferences drawn. Some of these ideas are included in Chapter 2; we should consider whether the points can be made more clearly.

4. I suggest some discussion could be added about specific judgments about specific exposure models that are then applied to inference about NO2/NOx effects, most likely in the context of the specific studies used to judge causality. I suggest that it would be appropriate to give higher weight to studies that do a better job taking into account the street network in the inference (note that in some applications there may be technical reasons why obvious choices, such as LUR models, aren't always better; see Szpiro et al 2011 Epidemiology) and less weight to those that will miss it completely. This may be particularly important for NO2/NOx (vs. e.g. PM). Here are some suggestions:

    a. Models that rely only on the existing regulatory network (at least prior to the near-road monitoring network) may not adequately capture the increased exposure near roads due to too few monitors in the network that are sited near roads.
    b. Nearest monitor exposures (e.g. Miller et al 2007) may not reflect NOx exposures for many individuals (again depending on how the monitors are sited), thus potentially strongly affecting the ability of such studies to detect health effects if they indeed exist. It could be interesting to contrast the relative merits of nearest monitor exposure estimates for spatially heterogeneous NOx vs. the much more spatially smooth PM2.5.
    c. IDW exposure estimates (e.g. Lipsett 2011) may smooth over road networks too much, unless there is an extremely spatially dense monitoring network used. Again the ability to detect NOx effects may be extremely poor in such a situation.
    d. Dispersion models may only capture some sources of NOx. There could also be important systematic errors in dispersion models due to how key assumptions are made and implemented. This would increase the uncertainty of the findings from studies that rely on dispersion models as the estimates could be better or worse than one might anticipate if the

true exposures were known. Because some of the errors are likely to be systematic with dispersion models, it may be more difficult to characterize their direction.

5. I suggest that there be an expansion of the measurement error perspective for inference about health effects. This is discussed in Chapter 2 but based on my preliminary review some important distinctions do not come across clearly. Review of a recently published discussion paper should be added and its perspectives brought into the discussion (Szpiro & Paciorek, 2013 Environmetrics with discussion by Spiegelman, Thomas, Hodges, Peng). That paper focuses on cohort studies where the key source of exposure variation is spatial; this perspective needs to clearly be stated as part of the discussion. Of particular importance are the following concepts:

    a. Exposure predictions have measurement error that can be decomposed into Berkson-like and classical-like components. The Berkson-like component comes from the prediction not capturing all the variation of the true exposure. The classical-like component comes from the uncertainty in the estimates in the exposure model. Neither component is true Berkson or classical (thus the "-like" terminiology) because the information used to derive the predictions is shared across all subjects. (There is mention of Berkson- and classical-like errors in Chapter 2, but so far in my quick review I did not see these terms defined in the document. My review of the concepts here is intended to make sure the understanding of these concepts comes across clearly.)
    b. The monitor and subject locations should be compatible, i.e. come from the same underlying location distribution.
    c. Spatially structured adjustment variables in the health model should be included in the exposure model.

6. In Chapter 2 I think the target exposure for inference should be defined in the context of the exposure measurement error discussion. Is it and should it always be total personal exposure? Or should it be personal exposure to ambient-source pollutants? When is it appropriate to consider ambient concentration as the target exposure for inference? In measurement error research, there are a whole host of issues in understanding the role of measurement error when the target exposure is ambient concentration. It will be important to consider those, and to address them distinctly from the issues that arise when the target exposure for inference is total personal exposure.

In Chapter 5 there seems to be an artificial distinction in the document between "measured" NO2 and modeled NO2. I would dispute that an estimate of NO2 based on IDW or nearest monitor is any more "measured" than an estimate based on LUR.