Preliminary Comments from Members of the CASAC Sulfur Oxides Panel on EPA's Risk and Exposure Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides (External Review Draft – August 2017)

Received as of 09-13-17

Dr. John Balmes 2
Dr. Judith Chow 4
Dr. Aaron Cohen 7
Dr. Alison Cullen 9
Dr. Steven Hanna 12
Dr. Jack Harkema 19
Dr. Elizabeth A. (Lianne) Sheppard 21
Dr. Frank E. Speizer 24
Dr. James Ultman 26
Dr. Ronald Wyzga 27

Dr. John Balmes

Introduction and Background for the Risk and Exposure Assessment (Chapter 1)

1. Is the introductory and background material, including that pertaining to previous SO2 exposure/risk assessments, clearly communicated and appropriately characterized?

Yes

Conceptual Model and Overview of Assessment Approach (Chapter 2)

2. Does the conceptual model summarized in section 2.1 adequately and appropriately summarize the key aspects of the conceptual model for the assessment?

Yes

3. Does the overview in section 2.2 clearly communicate key aspects of the approach implemented for this assessment?

Yes

Population Exposure and Risk (Chapter 4)

8. Is the presentation of, and approaches used for, key aspects of the exposure modeling, including those listed below, technically sound and clearly communicated?

a. Representation of simulated at-risk populations (section 4.1).

Yes

b. Estimation of elevated ventilation rate (section 4.1.4.4).

Yes

c. Representation of microenvironments (section 4.2).

Yes

d. Derivation of the exposure-response functions (section 4.5.2).

Yes

Exposure and Risk Estimates (Chapter 5)

9. This chapter is intended to be a concise summary of exposure and risk estimates, with interpretation with regard to implications in this review largely being done in the PA. Is the information technically sound, appropriately summarized and clearly communicated?

Yes

Characterization of Uncertainty and Representation of Variability (Chapter 6)

10. What are the views of the Panel regarding the technical appropriateness of the assessment of uncertainty and variability, and the clarity in presentation?

a. To what extent has variability adequately been described and appropriately represented (section 6.1)?

The variability has been adequately described.

b. To what extent have sources of uncertainty been identified and their implications for the risk characterization been assessed (section 6.2)?

Sources of uncertainty have been adequately identified and their implications for risk characterization have been reasonably assessed.

Dr. Judith Chow

Chapter 3: Ambient Air Concentrations

Chapter 3 is well-written and includes study area characteristics, air quality modeling input, a rationale to select air quality model receptors for exposure modeling, adjustments to design values, and estimation of continuous 5-minute data. Following are responses to the four charge questions for Chapter 3:

4. Does the Panel find the description of the three study areas and their key aspects (section 3.3) to be clear and technically appropriate?

The five criteria used to select individual study areas seem reasonable. They consider: 1) design values near the SO_2 standard (75 ppb); 2) one or more air quality monitors reporting 5-minute SO_2 data for the study period; 3) availability of sufficient air quality modeling data; 4) population >100,000 people; and 5) significant and diverse SO_2 emissions sources. It is good that the candidate study areas cover large geographic regions (i.e., New England, Ohio River Valley, and Midwest) and contain a variety of SO_2 emissions sources (e.g., electricity generating units [EGUs], secondary lead smelter, and petroleum refinery). However, AERMOD modeling uses the 2011 National Emissions Inventory (NEI) and compares SO_2 data from the 2011-2013 period rather than the most current data available. Assuming no major changes in emission sources over the past few years, comparison of 2011-2013 SO_2 data with the most recent measurements (2014-2016 or 2013-2015) should be made to confirm that there are few changes or reductions in SO_2 concentrations over recent years and to justify the use of 2011-2013 data.

The three study areas differ from those presented in the six focus areas of the Second Draft ISA (U.S. EPA, December 2016). The six locations evaluated in the Second Draft ISA include: Cleveland, OH, Pittsburgh, PA, New York, NY, St. Louis, MO, Houston, TX, and Gila County, AZ. These locations vary significantly from the three study areas selected for air quality exposure and monitoring in the REA. In addition, the criteria used to select the six focus areas include: 1) relevant current health studies; 2) existence of four or more monitoring sites located within the area's boundaries; and 3) the presence of several diverse SO_2 sources (U.S. EPA, December 2016). Relevant current health studies should be important criteria for consideration. The REA needs to justify the selection of the three study areas which differ from the six focus areas and document relevant health studies for the three study areas.

The approaches used to define the exposure modeling receptors within the air quality modeling domain (Section 3.3) are reasonable. Although the Fall River (MA) area uses a fixed 500 m grid, the Indianapolis (ID) and Tulsa (OH) study areas have receptor grids as low as 100, 250, and 500 m near major emitters and at various spatial scales. The 1400-1900 air quality model receptors

within 10 km of the major sources for each study area can represent population exposure adequately.

5. Does the Panel find the description of the air quality modeling done to estimate the spatial variation in 1-hour concentrations (section 3.2) to be technically sound and clearly communicated?

Section 3.2 documents detailed model inputs (e.g., meteorological measurements, surface characteristics and land use, emission sources, terrain, and air quality receptor locations). The designation of certain ambient monitors as background sites for each study area needs to be clarified. The intentsion is to remove potential impacts from local sources or represent background or boundary conditions for air quality modeling. Therefore, certain hours when wind directions indicate contribution from local sources are excluded. However, the representativeness of metrological data used to determine the number of hours to be excluded from the calculation of background concentrations needs to be justified. Table 3-7 (Page 3-13) shows hourly background concentrations of 16-18 ppb during summer at the Fall River study area; the adequacy of using these high background concentrations needs to be justified.

Among the three study areas, Indianapolis has the most complex set of point sources (Table 3-6, Page 3-9 and Figure 3-2, Page 3-18); the approaches taken to estimate hourly background concentrations stratified by season need to be documented.

6. To simulate air quality just meeting the current standard, we have adjusted model predicted 1-hour SO2 concentrations using a proportional approach focusing on the primary emissions source in each area to reduce the modeled concentrations at the highest air quality receptor to meet the current standard (section 3.4). Considering the goal of the analyses it to provide a characterization of air quality conditions that just meet the current standard and considering the associated uncertainties, what are the Panel's views on this approach?

Although the steps taken for air quality adjustment seem logical, Table 3-8 (Page 3-16) shows that the modeled air quality receptor maximum design value for Indianapolis is 311 ppb with a proportional adjustment factor of 4.21. The uncertainties of using high adjustment factors to estimate exposure need to be addressed.

7. A few approaches were used to extend the existing ambient air monitoring data to reflect temporal patterns in the study area (section 3.5). Does the Panel find the approaches used below to be technically sound and clearly communicated?

a. Data substitution approach for missing 1-hour, 5-minute maximum, or 5-minute continuous ambient air monitor concentrations (section 3.5.1).

There doesn't appear to be a better alternative.

b. Estimating pattern of within-hour 5-minute continuous concentrations where 1-hour average and 5-minute maximum are known (section 3.5.2).

Again, there are no other alternatives. As many monitoring sites have started to acquire continuous 5-minute SO_2 data since 2010, EPA is encouraged to require states to start reporting twelve of the 5-minute measurements within each hour, consistent with the new SO_2 monitoring guidelines stated in the Second Draft ISA (U.S. EPA, 2016).

c. Combining pattern of continuous 5-minute concentrations within each hour from monitors in or near the study area with the modeled 1-hour concentrations (section 3.5.3).

The example given in Section 3.5.2 illustrates the applicability of using the continuous 5-minute data from 2011 and 2012 at the Fall River study area to estimate 5-minute data for 2013; it confirms the assumption of log-normal distributions, categorized by SO_2 data 1-hour average with their peak mean ratios. However, the high concentrations found in Fall River represent the best case scenario, assuming climatology didn't change from 2011-2012 to 2013. This doesn't necessarily represent the Indianapolis case. As the Indianapolis study area did not have any continuous 5-minute monitor data, the surrogate monitor from Wayne County (Detroit, MI) was selected based on geographic region and similar design value. Comparisons should be made to demonstrate the deviations of estimated 5-minute data in worst case scenarios and representativeness of meteorological conditions reported by the surrogate monitor.

References

U.S. EPA (2016). Integrated Science Assessment for Sulfur Oxides—Health Criteria, Second External Review Draft. EPA/600/R-16/351. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711.

U.S. EPA (2011) National Emissions Inventory Report. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711. https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-data

Dr. Aaron Cohen

Exposure and Risk Estimates (Chapter 5)

9. This chapter is intended to be a concise summary of exposure and risk estimates, with interpretation with regard to implications in this review largely being done in the PA. Does the Panel find the information here to be technically sound, appropriately summarized and clearly communicated?

The information appears technically sound and has been, for the most part, appropriately summarized. It could, however be more clearly communicated. See my specific comments below.

Specific Comments

Page 5-1, lines 4-5 - I think that Figure 2.2 is intended to provide a visual summary of this process. If so, it should be called out here, or, perhaps even repeated so readers can refresh their memories about how the estimates of exposure and risk were derived....

Page 5-2, lines 11-15 - So the basic unit of estimation and analysis was the census block? If so, it would help to state this more clearly and explicitly...

Page 5-5, lines 10-16 - Provide some quantitative results to illustrate how sensitive the estimates were...

Page 5-5, lines 29-31 - Where could readers find the evidence for this in the REA?

Page 5-6, lines 10-11 - Where in the REA might readers find the evidence for this? Perhaps add time-activity summaries to Table 5-2.

Page 5-7, line 10 – Suggest changing "…occurrences focused in the Fall River study area..." to "occurrences largely limited to the Fall River area…"

Page 5-7, line 16 – "DVs" Should either say "DV" or "design value." Pick one term and use it throughout...

Page 5-9, line 10 - Are there relevant differences in source-specific contributions among the 3 areas?

Page 5-10, lines 14-23 - I suggest a simple declarative sentence(s), perhaps at the beginning of this section that summarizes this phenomenon...

1 Page 5-14, lines 27-28 - These are not summarized in Section 5. Should they be?

Dr. Alison Cullen

Chapter 4 Population Exposure and Risk

Question 8. Does the Panel find the presentation of, and approaches used for, key aspects of the exposure modeling, including those listed below, to be technically sound and clearly communicated?

a. Representation of simulated at-risk populations (section 4.1)

The representation of at-risk populations is approached in a technically sound manner and is clearly communicated. There are several issues related to spatial variability associated with geographic location that could be clarified.

- Although the importance of age and gender distributions of employment is clearly communicated, it should be stated clearly here whether information about spatial differences in the underlying age distribution of the population in the three study locations is included in this analysis. The distribution of age differs by location with family size and other population parameters as referenced in Table 6-2 of the REA draft but which particular co-variances are included in the simulation would benefit from more detail and a clear statement as mentioned in the comments about the PA.
- Energy expenditure by different individuals is modeled using appropriate and current literature on resting metabolic rate. Does the spatial profile of temperature and season affect resting metabolic rate such that differences in the three study locations related to individuals regulating personal temperature in the face of different climatic averages should be represented? If this co-variance is included in the analysis it should be references and highlighted in section 4.1.4.3. If not, a brief statement that this variability is dominated by other contributors to overall variance, if this is the case, would be helpful.
- Given that quite a bit of representation of correlation and co-variance was carried out for this analysis, it would illuminating to see a sensitivity analysis about which of these correlations and co-variance inclusions actually had an impact on the results. A comparison of the analytic results with and without the co-varying relationships accounted for would be valuable for the SO2 NAAQS process and possibly that of other air contaminants.

b. Estimation of elevated ventilation rate (section 4.1.4.4)

The approach to estimation of elevated ventilation rate across populations and conditions is clearly communicated in the REA with both inter- and intra-personal variability represented.

c. Representation of microenvironments (section 4.2)

The section on microenvironments is sound and based on a solid literature. The majority of peak exposures to SO2 occur outside and rely on a microenvironmental approach for the exposure simulation. The importance of AER and the role of air conditioning is well explained. One question pertains to a potential correlation or interaction between socioeconomic diversity and the presence of air conditioning (and implications for AER). Table 6-2 of the REA has general information about the co-variances included however it is not clear about the status of this one in the analysis.

Also, regarding human activity patterns (in a previous section, i.e., section 4.1.5) one question about the use of less than a third of the CHAD data remains unanswered. The statement is made (and accurately) that the other two thirds of the CHAD data does not include a breakdown of time spent indoors and outdoors by the participants in ATUS (the American Time Use Survey). This is an important ratio, but could this ratio not be developed based on the CHAD data for which the indoor/outdoor information is available and then applied to the other two thirds of the dataset? Given the amount of data that is unusable on this basis it is worth at least a comment regarding why such an estimated ratio and assumption is not applied (especially in light of the many assumptions that are necessary and included in the analysis as it stands).

d. Derivation of the exposure-response functions (section 4.5.2)

The development of a probit model for lung function risk as an exposure-response function is well reasoned. As referenced, in the earlier ISA second draft a doubling of sRaw, or increase of 100%, is defined as a moderate lung function decrement. The inclusion of an increase of 200% is added to represent a more severe lung function decrement.

The top panel in Figure 4-1 shows the probit form fit to the data points assuming sRaw greater than or equal to 100% (doubling) and illustrates well the concerning issue that the data points reflect a great deal of variability at the lowest dose range (200 to 300 ppb), the range which is closest to the levels of concern that drive the standard. In fact between 250 and 300 ppb all six measured data points are associated with a response that falls outside of the 5th and 95th percentile envelope around the probit fit, including some where much higher fractions of the studied population responded. Additionally these much higher fractions of the population represent a great number of individuals. The bottom panel with sRaw greater than or equal to 200% fares somewhat better in this regard.

On page 4-25 line 26 – 28 an illustrative example is used to explain the interpretation of the information gleaned from the probit model. The example refers to binning the exposure and representing the 10-20 ppb bin with the response level associated with its midpoint (15 ppb) as obtained from the probit. The actual value for the estimated response associated with this point is

omitted from the text; however, its inclusion would complete this example and improve its clarity substantially.

Dr. Steven Hanna

Note that my expertise is primarily in atmospheric transport and dispersion modeling, analysis of observed concentrations, and uncertainty and variability, and my comments focus on those areas. I was asked to comment on the areas related to Chapter 3 "Ambient Air Concentrations" and associated Appendices A, C and D. In addition, I comment on Chapter 6 "Variability Analysis and Uncertainty Characterization".

Charge questions for the Panel's consideration are presented below for Chapters 1, 2, 3 and 6. I assume that Appendices A through D fall under Chapter 3 and include those appendices in my comments.

Introduction and Background for the Risk and Exposure Assessment (Chapter 1)

Charge Question 1. Does the Panel find the introductory and background material, including that pertaining to previous SO2 exposure/risk assessments, to be clearly communicated and appropriately characterized?

There is a tendency to assume that the reader is already familiar with the previous reviews of the NAAQS for SOx and the various key issues. Since I joined this panel only about three years ago, much of the previous decades of work has been sometimes puzzling to me, in the sense that it does not conform to my research experiences. Surely it is possible to explain the key facts in a succinct way that is understandable to most readers. I have a hard time finding exactly where the health end-points or criteria are clearly defined in this chapter. Similarly, it would help if the new information covered in the current REA could be more clearly explained. What is new and different here?

I find that there are many places in the procedure where subjective or arbitrary choices are made, with insufficient justification in the written material. There should be more discussions of pros and cons and why specific choices were made.

My attempt at a simple explanation is: Based on SO2 health studies, we would really prefer a 5 minute standard. However, the EPA's dispersion model, AERMOD, does not produce concentrations averaged over less than 1 hour. And monitors (prior to about 2010) produced only one-hour averages. So the 2010 SO2 standard was based on 1 hour averages, but using statistical relations between available 5 minute and one hour peak SO2 concentrations to protect at a 5 minute level. To help us better understand 5 minute SO2 averages, monitors were converted so that they could measure 5 minute averages. In the current report, these additional 5 minute data are further analyzed to improve knowledge of their relation to the one hour averages. Three geographic domains with large SO2 sources are used as test cases, where AERMOD is run, but only to produce relative spatial variations in SO2 concentrations over the domains.

Conceptual Model and Overview of Assessment Approach (Chapter 2)

Charge Question 2. Does the Panel find the conceptual model summarized in section 2.1 to adequately and appropriately summarize the key aspects of the conceptual model for the assessment?

Although the conceptual model has to follow an obvious chronological sequence, with health effects and risk calculated at the end, it would be helpful to first state the precise definition of the "health endpoints" and "risk metrics". This is needed to better understand the rationale for the emissions modeling, the use of observed and meteorological data at various averaging times, and the air modeling approach.

The "sources" subsection 2.1.1 gives a comprehensive list of source types. Please justify why only the few major point sources of SO2 are modeled in this report. The three test domains also have significant emissions from traffic and area sources (in which many smaller point sources are lumped).

Charge Question 3. Does the overview in section 2.2 clearly communicate key aspects of the approach implemented for this assessment?

The subsection is only 2 ½ pages, and 2/3 page is occupied by Figure 2-2 with the approach's flow chart. I note that this figure is similar to Figure 2-1, at the beginning of chapter 2. Can these two figures be combined?

How was it decided that there would be three study areas? Why not 2, 5, or 10?

Paragraph at top of p 2-7 – I'm not sure that it is realized that much of the variability in 5-minute concentrations within a given hour is due to inherent variability in atmospheric turbulent eddies. The best example is a sunny afternoon when there are convective eddies with time periods of 5-10 minutes that result in a plume looping or flopping around with that period close to the source (within a few km). Footnote 9 at bottom of page – Clarify that these are EPA models that are being discussed. Many other agency and country models can simulate concentrations at small averaging times (1 -10 minutes) and are applicable to large SO2 sources. SCICHEM is an example.

It is good that the risk metrics are better described on p 2-8 lines 6-11. It would also help if EPA could state what they consider to be "acceptance criteria" (standards?) for these metrics.

Ambient Air Concentrations (Chapter 3)

Charge Question 4. Does the Panel find the description of the three study areas and their key aspects (section 3.3) to be clear and technically appropriate?

I agree that the three study areas (Fall River, Indianapolis, and Tulsa) are reasonable choices. As I said earlier, there should be reasoning given for why only three areas were chosen (budget, time, staff, ??). In most subsections, the justification and the details could be improved. For example, for this type of area, standard AERMOD modeling would include many more sources, including traffic and area sources; yet this exercise only modeled the few major SO2 sources. Details such as maximum receptor distances from the source should be justified.

One criterion for choice of study area was that there should be one or more monitors reporting 5-minute SO2 concentrations. Yet representativeness does not seem to be a major concern, since the monitor is often considerable distance from the source. It is well-known that larger fluctuations in concentration for 5-minute averages can occur closer to the source. So the observed 5-minute variability at the monitor may be quite different from that at the location where the design concentration occurs.

Charge Question 5. Does the Panel find the description of the air quality modeling done to estimate the spatial variation in 1-hour concentrations (section 3.2) to be technically sound and clearly communicated?

My comments also cover Appendices A, C, and D, which support section 3.2. In most cases, the AERMOD runs made use of input data and scenarios run by others on these domains. Inputs of meteorology, terrain, land use etc. are standard and are, in most cases, already approved by local agencies.

p 3-5 – Meteorological inputs – These are standard, but they suffer from the usual representativeness problem. The surface winds at a given hour at the reporting site may be different from that 10 km away at the pollution source. At all three sites, the upper air data are from sites 100 to 200 km away. It should be mentioned that this lack of representativeness can contribute to the uncertainty and variability of the results. Most of the text on p 3-5 describes an empirical method used by the authors to mitigate the “problem” of too many calm winds reported in the hourly data; I can’t follow the method that is described and wish it could be better explained.

p 3-6, lines 13-16 – Better justify the use of older land cover data rather than just saying “is not expected to have a significant effect”.

Background concentrations section 3.2.4 – Please reword this section so that it is clearer. Because you are not modeling sources other than the very few large point sources, and you have chosen populated areas, there are many moderate point sources, line sources, and area sources that must be accounted for somehow in your approach. Instead, you appear to be relying on observations from a few monitors, and considering “background” to be the readings from time periods when the wind is not blowing the plume from the AERMOD-modeled large point sources to the monitor. It might help your explanation to include a diagram showing the monitor location, the point source location, and the wind sectors.

Section 2.5 and Appendix D – Hourly concentrations at the air quality model receptors. The main text has only seven lines, so most of my comments are on Appendix D, where the modeled and monitored concentrations are compared. Many Q-Q plots are included and some quantitative performance measures. I am most interested in the comparisons of design values – At Fall River (Table D-1), AERMOD is about 20 % low at the one monitor. At Indianapolis (Table D-4), the model is about 40 % high at monitor 57, 30% low at monitor 73, and 20 % low at monitor 78. At Tulsa (Table D-5), the model is 30 % low at monitor 75, 30 % high at monitor 35, and almost right-on for monitor 27. I conclude that the model is agreeing with these observed design values within plus and minus 30 or 40 %. This is typical. However, as said before, there are only eight monitors operating across the three areas, and that isn't enough to capture the "real" observed maximum concentrations around the major point sources. Also, we do not know if the observed and modeled maxima are occurring during different times of day or wind speeds (the design value doesn't care, though!)

Charge Question 6. To simulate air quality just meeting the current standard, we have adjusted model predicted 1-hour SO2 concentrations using a proportional approach focusing on the primary emissions source in each area to reduce the modeled concentrations at the highest air quality receptor to meet the current standard (section 3.4). Considering the goal of the analyses it to provide a characterization of air quality conditions that just meet the current standard and considering the associated uncertainties, what are the Panel's views on this approach?

I can see how you are doing this but I do not see why. Perhaps the why was discussed several years ago, the last time the SO2 standard was revised. Thinking that I might be missing something obvious, I asked several of my colleagues whose job is to run AERMOD on a daily basis, and they could not suggest why. However, they are mostly running permitting exercises, and maybe this procedure is unique to the follow-on exposure and risk model, APEX. Could the authors of the current report explain why this scaling to the design value is done?

There is an implicit assumption that the concentration distributions in these metropolitan areas are dominated by one or two big point sources. Also, you are ignoring the fact that application of the scaling method implies that the emissions are going up or down, and this will also affect the plume rise. Plume rise in turn affects downwind ground level concentrations. So the process is actually nonlinear.

Charge Question 7. A few approaches were used to extend the existing ambient air monitoring data to reflect temporal patterns in the study area (section 3.5). Does the Panel find the approaches used below to be technically sound and clearly communicated?

As commented earlier, I have some difficulty understanding the various methods and subjective assumptions in this report. A few key details are left out. Regarding "technically-sound", I would say that the method is more "statistically-sound" than "technically-sound". Concentrations vary in the atmosphere due mainly to variations in wind and turbulence. Scientific models and

formulations for variances, covariances, and time and space scales are well-documented in the boundary-layer literature. AERMOD's formulations partially account for these factors (but are not evident to the routine modeler).

In reality, the 5-minute averages will vary more near the large point source and during periods with significant lateral meandering or large convective eddies. Thus there is a problem in that the monitors providing the 5-minute data are not close to the source, and will likely show less variability (i.e., smaller peak-to-mean ratios).

Charge Question 7a. Data substitution approach for missing 1-hour, 5-minute maximum, or 5-minute continuous ambient air monitor concentrations (section 3.5.1).

The methods are similar to what is done in AERMET for missing meteorological data. In most case, though, the data substitution methods will underestimate the actual variability. In one case described in this section, data are used from a surrogate site in Detroit, and I would have to look at more details to decide whether this is acceptable. In addition, as stated earlier, I think that there is an assumption of random variability that is made that may not be valid at times when the turbulence scales (e.g., convective eddies and mesoscale meanders) are causing correlations.

Charge Question 7b. Estimating pattern of within-hour 5-minute continuous concentrations where 1-hour average and 5-minute maximum are known (section 3.5.2).

This is probably fine as long as the monitoring site is representative of the location where the 5 minute variations are to be used. At the end of the subsection, six Q-Q plots and a table are shown where measured and estimated variables are compared, and more details are needed to aid my understanding. It would help to have better captions and axis labels. For example, in the upper left plot, why don't all the points fall along the line of perfect agreement? In the bottom plots, there is a lot of scatter, with scatter larger than the observed value at C < 50 ppb. Why is that? There is no explanation in the text.

Charge Question 7c. Combining pattern of continuous 5-minute concentrations within each hour from monitors in or near the study area with the modeled 1-hour concentrations (section 3.5.3).

Here too there may be a representativeness problem due to the several km separation between the monitor and the location of the model prediction. To check this out, you could use the measured data from, say, Indianapolis, where there were four monitors, and correlate the 5 minute data from one site with another.

p 3-29, lines 22-32 – I can follow most of this, until I get to the ranked concentrations part. Why not just combine the modeled 1-hr average time series of concentrations with the observed 5-minute averaged concentrations (expressed as differences from the observed 1-hour average) for each hour. This seems very straightforward. However, if you have scaled the modeled

concentrations with the design value, you should do the same with the observations (apples to apples). In lines 25-26 you assume that the spatial gradients are unchanged; however as I state above, the turbulent variability that exists in time also exists in space, and you are not accounting for the spatial turbulence (SCIPUFF does this automatically).

pp 3-30 to end of chapter – Better explanations are needed of these comparisons. Be precise in text definitions and in table captions and row and column headings

Characterization of Uncertainty and Representation of Variability (Chapter 6)

Charge Question 10. What are the views of the Panel regarding the technical appropriateness of the assessment of uncertainty and variability, and the clarity in presentation?

This approach seems to be similar to what the EPA has applied in previous reports, modified to account for the 5-minute average concentration assessments. The general discussion is fine; my main comments concern specific sources of uncertainty, focused on meteorological data, AERMOD modeled 1 hour concentrations, and ambient monitor 5-minute concentrations.

Charge Question 10a. To what extent has variability adequately been described and appropriately represented (section 6.1)?

The surface meteorological data site is as much as 20 km from the edges of the modeling domain. This raises issues about representativeness. Even annual wind roses can vary significantly across this distance. In our model evaluations, we often test several possible wind sites and can see significant differences, especially in the direction with the maximum concentration. The report does not discuss this. Also, the upper sir site is 100-200 km away, and wind vectors and stability and mixing depths can vary quite a bit over that distance.

The AERMOD modeled concentration fields are acknowledged in its technical document to represent an ensemble average, and, on any individual hour in the ensemble, there can be variations.

As for the surface met site, the 5-min average concentration monitoring sites may not be representative of the location where AERMOD is calculating the concentration.

Charge Question 10b. To what extent have sources of uncertainty been identified and their implications for the risk characterization been assessed (section 6.2)?

I'm looking at Table 6-3. Under AERMOD meteorological data, I would say that the wind direction uncertainty (unrepresentativeness) could have a high influence on the total uncertainty.

The actual wind at the plume location might blow more often towards a population center, while the observed wind (at a NWS location 10 km away) might not.

Under AERMOD algorithms, it should be recognized that EPA evaluations mainly consider only observations and predictions unpaired in time and space. This is why Q-Q plots are favored by EPA. If you pair the data in time and space, there is a lot of scatter (typically a factor of two), mostly due to wind direction differences.

Under ambient monitor 5-minute averaged concentrations, I believe that the same nonrepresentative problem exists as for the surface meteorological data. We know that concentration variability is larger closer to the source. Also, there is a known mesoscale spatial and temporal time scale that is not being accounted for. That is, if the concentration is relatively large during one 5-minute period, it is likely to be relatively large during the next period. Typical observed mesoscale turbulence time scales are 5 to 10 minutes and space scales a few km.

Dr. Jack Harkema

Chapter 1

I find the introductory and background material, including that pertaining to the previous SO_2 exposure/risk assessments, to be adequately communicated, and appropriately characterized. The authors, for the most part, have rightly avoided using excessive technical jargon and have produced a clearly and concisely written introductory chapter.

The authors should consider, however, revising (shortening or deleting) the first three paragraphs of this Chapter since much of this information is covered (sometimes verbatim) in 1.1 (Background) and the remaining sections of this chapter.

Early in this chapter the authors should state briefly, but clearly, the justification/rationale for using quantitative analyses of SO_2 for risk and exposure assessments for SOx (see p.1-6, lines 29-31 in PA document as a suitable statement in this regard). This explanation could be reiterated in 2.1.1 as well.

p. 1-2, line 8. Consider inserting “adverse” after “. . . identifiable” and prior to “effects on public health . . .”

p. 1-2, line 25. The second half of this sentence could be clearer. The authors should consider revising (see 1-6, lines 32-36 in PA as a more clearly written statement).

Chapter 2

I find the conceptual model summarized in 2.1 to be very useful and effectively illustrated with Figure 2-1. The overview in section 2.2 does clearly communicate the key aspects of the approach used for this assessment.

Authors could consider renaming the title of the population box in Figure 2-1 to “Children and Adults with Asthma” or “Exercising Adults and Children with Asthma.”

Authors could also consider revising the Lung Function Risk box to read “. . . experiencing moderate or <u>severe</u> lung function <u>decrements</u> . . .”

Furthermore, in the exposure section of Figure 2-1, a box entitled “Pulmonary Airways” or “Lung” could be inserted between the Inhalation box and the People with Asthma box, and subcategorized as “Dose.”

p. 2-2, line 2. Consider inserting “direct” prior to “emissions of SO_2 . . .”

p. 2-2, line 9. The authors should provide an example of "other industrial facilities."

p. 2-6, lines 28-29. A brief explanation for why the authors used "air quality conditions simulated to just meet the current standard" should be provided.

Dr. Elizabeth A. (Lianne) Sheppard

General comments

Overall I found the document to be clear and appropriate for its goals. There was sufficient detail and cross-referencing to support the analyses and background presented. The study of three areas is preferable to fewer.

CQ 10: What are the views of the Panel regarding the technical appropriateness of the assessment of uncertainty and variability, and the clarity in presentation?

I found the presentation to be appropriate, clear and well-organized.

a. To what extent has variability adequately been described and appropriately represented (section 6.1)?

I think it is appropriate for the staff to use observed variability in the input data when these data are available and sufficiently representative. I think Tables 6-1 and 6-2 have a nice layout and provide good summaries. The lists seem comprehensive to me and no other factors come immediately to mind.

A few comments about the sources of variability:

1. Table 6-1: Spatial variability for the 5-minute concentrations is limited by the small number of monitors available and their placement. I'm not convinced that using the nearest monitor is the best way to represent spatial variability in these data. I believe this is an important limitation in adequately characterizing the spatial variability of 5-minute SO2.
2. Table 6-1: Clarify the E-R function regression estimates entry. The table implies that variation is captured from the upper and lower CIs. However, the upper CI estimate was only used in an uncertainty analysis. In my initial reading of this table, I assumed that estimates were generated at an individual level following the distribution with central tendency estimated by the probit model and variation in that by the CI estimates. This appears to be incorrect.
3. Table 6-1: The table description of the exposure bins for the E-R function is unclear to me. If it is not easily expanded to clarify, perhaps an alternative solution would be to reference the section in the document where this is outlined.
4. Table 6-2: I think I agree with the statement that between-person correlations may not be important to take into account. However I found the rationale confusing as shared population behaviors would tend to increase correlations rather than decrease them.

5. Table 6-2: It is good to acknowledge the lack of data for the correlation between activities and microenvironment parameters as the reason it isn't accounted for. I would not have expected this to be important, but the example suggests why it could be.

I have a few quibbles with the presentation:

1. I don't think the definition of correlation should be changed for purposes of this document. In statistics, correlation describes the strength of a linear relationship. If non-linear relationships are also being considered, then in my opinion they should not be referred to as correlations. I suggest using "relationship" instead of "correlation" and then indicating that relationships include both correlations (i.e. linear relationships) and other nonlinear relationships. (See e.g. the footnote to Table 6-2)
2. When the variation depends on more than one variable, they have a joint distribution. One might use conditional distributions to define joint distributions. In places the text lacks clarity about the concept of joint vs. conditional distributions w.r.t. to the intent. (e.g. p 6-3 line 2, use of the term "joint conditional variables" in Table 6-2)

b. To what extent have sources of uncertainty been identified and their implications for the risk characterization been assessed (section 6.2)?

Overall I think the uncertainty characterization is thorough and well done. Table 6-3 has a nice structure and seems comprehensive. Collectively we should continue to consider whether there are additional topics that should be added.

Is there any reason to consider a function other than the probit for the E-R function? (e.g., Section 6.2.3)

Most of my suggestions for improvements in this section have to do with presentation:

1. When possible I think ordering should be consistent throughout the section. For instance, the sections for sensitivity analyses in the table are presented out of sequential order. In the results tables, sometimes the primary analysis results are shown first, but not always. Sometimes there appears to be a clear rationale for the presentation ordering of the primary vs. sensitivity analyses; other times it is not apparent why there is inconsistency across results presentations.
2. I think every table (in e.g. a footnote) should make it clear which results are the primary analysis and which are the sensitivity analysis. This will foster clarity and eliminate the need for readers to look these up in other parts of the document.
3. Typically the sensitivity analyses don't show much impact. However, when there is an impact, even a very small one, the direction of that impact is fairly consistently larger

across conditions. I think this should be stated directly and not left to the reader to ferret out. While I agree with the general interpretations offered that the sensitivity results don't alter the conclusions, I still think the reporting should be more transparent about the fairly consistent direction of the very small effects when they exist.

4. The comments in Table 6-3 should make it clear when they are describing sensitivity analysis results. This isn't immediately obvious for the estimation of continuous 5-minute concentrations category.

Dr. Frank E. Speizer

Chapter 3

Regard the use of years 2011-2013, there is a remarkable decrease in exposure levels between 2011 and 2012-3. This is particularly true for Fall River. I raise the question (and will not available to hear the answer) is there some that happened between these years to account for exposure differences? Alternatively, are we missing some data or did position of monitors change for the later years. Since the focus in Chapter 5 is on Fall River, this should be detailed here to avoid it coming as a question later.

Page 3-20, line 29: Simply for consistency sake should probably indicate the years for Indianapolis. (I know it is described in next paragraph).

Chapter 4

Page 4-27, Figure 4-1. Although the data are presented in Table 4-10, the use of % response in both is quite misleading. These are for most part frightfully small studies and as indicated in the table often the positive responses are seen in 1 or two people. This is what it is but I wonder if it would be better to characterize the findings “more qualitatively” in the Figures rather than relying on statistics.

Pages 4-27-4-30 sought of acts as a “teaser” for what to expect in Chapter 6. Not sure it could all have been summed up by a single sentence (see chapter 6) but will have to wait until Chapter 6 is reviewed to see just how redundant this is.

Chapter 5

The data and analyses discussed in this chapter is an excellent presentation of the 3 study areas considered. The tables and figures really tell the story well and should set up for Chapter 6 details. I congratulate staff for making this rather complex topic readable and informative. It appears to me to be technically sound, and appropriately summarized and communicated, but I am still concerned as to the direction of the exposure uncertainty created by the real differences between 2011 and the 2012-3 data.

Chapter 6

Page 6-5, table 6-1. In regard to the above, Microenvironmental spatial variability is mentioned as potentially varying within and between study areas. This raises a red flag to me as a potential explanation for the drastic variability between 2011 and 2012-3. A comment somewhere is needed.

Table 6-3 starting on page 6-8: Again, congrats to staff. I found this a most useful table to review in detail. I have to admit I was looking for details that would explain the discrepancy between data for 2011 vs other years, but reading for details (which I did not find) was impressive that the potential uncertainties have been characterized and considered.

Table B-1 Only suggestion so far that there was a difference in emissions in year 2011 vs other years with a doubling of output. Not as true for other cities. (At least a potential explanation for differences in exposure.

Dr. James Ultman

Chapter 5 – Exposure and Risk Assessment

Exposure and Risk Estimates (Chapter 5)

Charge Question 9. This chapter is intended to be a concise summary of exposure and risk estimates, with interpretation with regard to implications in this review largely being done in the PA. Does the Panel find the information here to be technically sound, appropriately summarized and clearly communicated?

Within the many uncertainties in the risk analysis pointed out in chapter 6, the general approach outlined in this chapter is technically sound.

Clarity is somewhat lacking, however. This is understandable given the multiple data sources and complex methodology that is used. Most troublesome to me was the lack of a clear definition of the subdivisions used in the study areas: “AERMOD special grid” versus “receptors” vs “census blocks” versus “census blocks.” Also of considerable importance throughout the document is “design value” which is never precisely defined or explained.

There are three important shortcomings in the analysis. First, children are assigned the same exposure-response curve as that derived from adult laboratory studies. I think that further elucidation of factors that may increase risk in children should be included (*e.g.*, children have developing lungs so that long-term effects of SOx exposure may be important, even though current research has not adequately addressed this). Second, the CHAD activity logs are not specific to individuals with asthma. And asthmatics, particularly children, may spend less time outdoors than nonasthmatics and exposure is greater outdoors than indoors. This factor leads to an overestimation of risk. Third, the REA is limited to three study areas. There should be more discussion of how these results may be viewed from a national public health perspective.

In figures 5-4 to 5-6, the overlay of DV and population density distributions, it is difficult to visually separate the two variables. It might be better to make a 2-d plot of this type with a single combined variable such as the product of DV with population density. In that case, the coincidence of large values of the two variables would appear as the darkest pixels and the coincidence of small values would appear as the lightest pixels.

Additional Comment

pg 4-27, fig 4-1: Given the large (but quantifiable) uncertainty of the E-R function, it would be productive to produce upper and lower bounding values for some of the risk estimates based on sRaw decrement.

Dr. Ronald Wyzga

Overall Comments:

The REA is an impressive and extensive document, the object of which is to estimate the potential risks of short-term SO_2 exposure to asthmatics. It concludes that the risks are very small to minimal at levels of exposure which would be associated with current and other possible standards. The amount of work carried out in this exercise is most impressive. Some changes or additions could be made to this work, but it is questionable whether these changes would modify the overall conclusions. I cite at least 3 areas where the analysis could have been altered/extended. In considering the prevalence of asthma in the three study areas used as examples in this document, adjustment was made for several factors, but apparently did not consider race, for which the prevalence varies. I don't offhand know the racial composition of the three areas studied, but unless they are radically different from the overall racial distribution for the regions used to estimate prevalence, I suspect the overall results would change little. It would be useful if the REA would briefly address this issue.

A second issue is the use of values for nearby geographic entities for the specific areas studies. I note, for example, that air exchange rates for New York were used to estimate those for Fall River (and Indianapolis, that also made use of data from Detroit) and those for Houston were used for Tulsa. While specific data may not be available, use of data form other locations can create uncertainties in the analyses that could impact results. For example, the housing stock in New York is very different from that in Fall River. Air conditioning prevalence between Boston and Fall River could differ as well. This uncertainty should be mentioned in the report, and its potential impact should be discussed.

The third issue is that the analysis considered the years 2011-2013. Several changes in emissions have occurred since that period, resulting in even lower estimates of SO_2 exposure. I note, for example, that the Brayton Point power plant in Fall River was closed in July 2017, the PSO Northeaster Power Station in Tulsa retired unit 4 in April 2016, the IPL- Harding Street Generating Station in Indianapolis stopped burning coal in February 2016, and the Citizens Thermal unit in Indianapolis is presently converting from coal to natural gas. All of these will significantly reduce SO_2 emissions and lead to even lower risk estimates. Although the analysis focused on the 2011-2013 time period for which extensive data are available, these changes in emissions should at least be footnoted.

Charge question 10: What are the views of the Panel regarding the technical appropriateness of the assessment of Uncertainty and variability, and the clarity in presentation?

a. *To what extent has variability been described and appropriately represented (Section 6.1)?*

b. To what extent have sources of uncertainty been identified and their implications for the risk characterization been assessed (Section 6.2)?

I am most concerned about any deviations from the assumed inputs into the model that would increase the potential risks. Given the magnitude of estimated risks, factors that reduce the estimated low risks are not of major interest in re-examining the SO_2 NAAQS. I am less concerned about accuracy of exposure/risk estimates than of potential underestimates.

There are, however, a couple of components in Table 6-1 that could be given more discussion: 1.) Asthma prevalence It is unclear to me why race was not considered in the estimates of asthma prevalence. I doubt that introduction of this factor would lead to significant increases in the overall risk estimates, but the potential omission of race as a predictor variable should be addressed in a brief discussion. 2.) Microenvironmental approach Limited available data made it necessary to utilize data from other locations to estimate risks for the three study areas. It is unclear to me how differences between the utilized data and the reality in the study communities would impact the results. I cite some examples above. I know the Fall River area reasonably well; the housing stock there is very different from that of New York; hence the estimates of air exchange rates for Fall River may be inaccurate. How inaccurate need they be to substantially change the risk estimates. Similarly the removal rates are tied to air conditioning prevalence. Boston data are used for Fall River. The difference in housing stock and socio-economic factors between Boston and Fall River might influence the overall results. How sensitive are the overall results to changes in these components?

<u>Table 6-3 characterizes key uncertainties and tries to estimate their influence.</u>

I have a few comments here:

Point Source Emissions and Profiles: Data from 2011 are used and are reasonably accurate; hence any inaccuracies in the 2011 data are likely to be small and little impact on the exposure/risk estimates; however there have several significant changes since 2011-13, and these would clearly impact the exposure/risk estimates. This should be noted somewhere in the document, if only as a footnote.

APEX Microenvironmental Concentrations: See above comments. I wonder if the “low” Knowledge-base uncertainty is correct for Indoor A/C Prevalence. This is also an element that is changing over time.

Specific Comments on document:

p. 3-9, Table 3-6: It should be clearly stated that these sources were for the period of study: 2011-2013. A footnote might indicate known changes that have occurred since 2013.

p. 3-20, l. 29: To be consistent the years should be included here as well.

p. 4-4, ll. 4-5: I have seen state-specific data. Is there any reason why regional data were used rather than state data?
ll. 10-18; Given the differences in prevalence among races, why is this not a consideration here?

P. 4-19, Table 4-7: Given the differences in housing stock and socio-economics between Boston and Fall River, is there any need to consider these in the use of Boston data for Fall River?