Responses to CASAC Questions on the PM PA from Consultant Dr. Sonja Sax

Questions from Dr. Tony Cox

https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD010919.pub2/abstract

Are the C-R models in Table C-1 appropriate, logically valid, and empirically well-validated, for answering the causal question of how changes in PM2.5 levels would change health risks?

To answer this first question generally, the methodology used in the Draft PA[1] (i.e., the BenMAP model) to estimate the human health risks (i.e., mortality) associated with the current and alternative PM NAAQS is not meant to establish causality, but to provide quantitative estimates of risks and provide information regarding the adequacy of the current NAAQS. The causality is determined by application of the EPA NAAQS framework in the ISA (discussed below).

As such, the risk assessment (RA) generally considers health endpoint for which EPA has already determined that the endpoint is "causal"or "likely causal" based on integration of the evidence in the ISA. It is noteworthy, however, that in the PA, EPA evaluated only mortality (all-cause, ischemic heart disease or IHD, and lung cancer) from long-term exposures and all-cause mortality from short-term exposures, as it noted that these health endpoints reflect "clear public health importance, the large number of epidemiologic studies available for consideration, and the broad availability of baseline incidence data (Draft PA, C-2)."[2] In fact, mortality does drive the health impacts and particularly the valuation estimates in the regulatory impact analyses. However, I would argue that it is still of interest to assess the health impacts for other endpoints that EPA deems are causally associated as this provides a wider perspective, and provides context. For example, one would expect to see more emergency room visits for cardiovascular endpoints, followed by hospital admissions, and lastly mortality attributable to PM exposures for this endpoint.

In the risk assessment, EPA seeks to quantify the health impacts from current conditions or modeled estimates for just meeting the PM NAAQS vs. zero levels of $PM_{2.5}$ (which is not realistic) compared to modeled conditions for achieving alternative NAAQS vs. a zero PM concentration. This approach yields the elevated "avoided" mortalities reported in the PA, which is misleading because they are not realistic (i.e., there is always some background PM exposure and a zero PM level is unattainable) and as EPA acknowledges the uncertainty below background PM levels is very large since there is little data. From the point of view of assessing the risk differences, it is unclear why EPA would not simply compare the current or just meeting the NAAQS with the alternative NAAQS scenarios as this difference is in essence what the risk assessment is addressing.

The estimates of health risks (i.e., mortality) or risk reductions (i.e. "avoided deaths") are, solely based on the associations observed in selected epidemiological studies (i.e., relative risks, or hazard ratios). Therefore, the risk assessment, which is used to answer the question of how changes in $PM_{2.5}$ levels change health risks, already assumes that there is a causal link (that is that the concentration-response

[1] Like the ISA, these analyses were previously presented as a separate risk assessment document that included more health outcomes and more sensitivity analyses. In addition, the document underwent its own review process and was therefore subject to more scrutiny than in the current evaluation.

[2] It is important to note also that the ISA has not been finalized and it is unclear whether there will be changes to the causal classifications that would also impact the risk assessment (e.g., for lung cancer mortality)

function or CRF can be used to quantify the change in health outcome per some change in air pollution concentration). For most studies, this is based on the relative risk (RR), which EPA converts to a beta health function for use in BenMAP. The methodology is described in the BenMAP manual. But in short for a log-linear relationship:

$$ß = \ln (RR)/ \Delta PM$$

The ΔPM is typically the difference between the baseline scenario and the control scenario. As noted above, EPA actually calculated the ΔPM as the difference between meeting the NAAQS and zero PM ambient concentrations and this is compared to the ΔPM of alternative NAAQS and zero PM ambient concentrations.

While EPA details the criteria for the selection of epidemiological studies to include in the risk assessment (see page C-3 of the Draft PA), it is noteworthy that there is no consideration of study quality in the selection of studies or any consideration for any potential biases that may be associated with these studies. That is, EPA appears to consider these to be high quality studies for which bias, including any confounding has been ruled out. Again, this analysis assumes a causal association has been established.

Another noteworthy aspect of the analysis is that EPA relies on studies that report National estimates, yet conducts an analysis for specific geographic locations. It would be more appropriate for EPA to apply specific CRFs for each location, as there is substantial heterogeneity in the CRFs both spatially (regional and city-specific estimates) as well as temporally (i.e., different seasons). Most studies would be able to provide regional or city specific CRFs for use in BenMAP. This would reduce the uncertainty of using a National estimate to estimate effects in a specific region or city.

An additional observation is that given a positive RR and the assumption of linearity, BenMAP will always estimate a reduction of risk with reduced PM concentrations, so even without doing the BenMAP analysis you would expect reductions in risk with reduced air pollutant concentrations (i.e., lower alternative NAAQS), but it is unclear how informative this really is. That is, since you have a relationship that is assumed to be linear and it does not matter where on the concentration-response curve you are, you will see a risk reduction with reduced air concentrations down to zero. Therefore, for this analysis to be useful, it is important to consider the uncertainty. That is, if you already assume a linear response, and a causal association, then the question becomes: are the observed reductions in risk statistically significantly different between the current and the alternative NAAQS standards?

While this deviates somewhat from the questions posed by Dr. Cox with regards causality, it is important to note that EPA does presents uncertainty associated with the risk estimates. This uncertainty, however, is based only on the statistical uncertainty of the CRF (the standard error), but does not include uncertainty associated with the modeled air quality estimates (i.e., estimates for attaining the current standard as well as alternative standards), uncertainties associated with the baseline health statistics or population estimates. All of these variables factor into the BenMAP estimates. Therefore, the uncertainty estimates presented by EPA only reflect uncertainty in one aspect of the model variables (i.e., the statistical imprecision of the model), although it is likely that this variable is the most uncertain. Still, if we just consider the CRF uncertainty (i.e., the 95% confidence intervals) presented by EPA in comparing the estimated mortality from just meeting the current standard to meeting alternative standards (both based on comparing to a zero PM concentration) we find that the mean mortality estimates for just meeting the current standard are within the range of uncertainty for meeting alternative

standards – that is that the reduction in risk is not likely to be statistically significantly different because of overlapping confidence intervals. This indicates that even though risk reductions are observed with alternative standards, these risk reductions are within the uncertainty bounds of the analyses. EPA needs to make this point in the PA, instead of focusing on only the point estimates. In addition, it is noteworthy that although EPA discusses qualitatively many additional sources of uncertainty (see pages C-77 to C-80) and many are rated by EPA as having a medium-high degree of uncertainty with the impact on both directions, the EPA does not discuss this uncertainty in the context of interpreting the risk results.

With regards to causality, as noted above, EPA determines causality as part of the evaluation presented in the ISA. As noted by several reviewers and in the ISA comments, the EPA causal framework could be significantly improved with the application of modern day systematic review practices. In particular, the current framework lacks a transparent and systematic approach to evaluation of study quality. Therefore, it is unclear how EPA selects and weights studies for determining and classifying outcomes with regards to causality. In addition, how EPA integrates the evidence is unclear. Again, the ISA process could be improved by providing greater clarity on how evidence from different scientific lines are integrated to make a final causality assessment. Given the very small relative risks from largely ecologic epidemiological studies with serious limitations, it is critical to assess all lines of evidence to determine causality.

In the Draft PA, there is some limited discussion of lines of evidence that EPA relies on other the epidemiological evidence. I would argue that the epidemiological evidence alone should not be the basis for assessing a causal association, and this clearly has implication for the risk assessment (i.e., if the epidemiological evidence cannot establish causality then it maybe should not be used to quantify health risks). On the one hand, EPA specifically states that evidence from controlled exposure studies and animal studies provide supportive evidence for the associations observed in epidemiological studies, but on the other hand EPA readily dismisses this evidence for providing information regarding the adequacy of the current NAAQS noting that because controlled exposure studies and animal studies were conducted at levels above the current NAAQS they were not informative. I disagree that this evidence is not informative, instead it is important to consider that if effects are not consistent or coherent at levels above the NAAQS in experimental studies, this may indicate that the observations in the epidemiological studies may be due to other variables that are unaccounted for.

Specific questions:

1. Are the beta coefficients in Table C-1 of the PA conceptually well defined?

In reviewing the PA, EPA does not explicitly define the beta coefficient that is used in BenMAP, or how it is derived, instead referring to prior evaluations and the BenMAP manual. EPA should include a definition of the beta coefficient and how it is derived for clarity, especially because both prior assessments and the BenMAP manual do not have specific information related to the studies that are considered in the PA. That is, prior assessments and information in the BenMAP manual include different epidemiology studies than those used in the PA evaluation. Similarly, EPA should define the concentration-response function and how it is used to derive the beta coefficient. In short, in the PA the beta coefficient is not well defined.

2. On the same topic of clear definitions, does the discussion of the BenMAP-CE beta coefficients in the PA and underlying documentation (described as typically representing "the percent change in a given

adverse health impact per unit of pollution") unambiguously specify which of the following concepts the coefficients represent?

The PA does not specifically define the beta coefficients or the concentration-response functions and how they are derived from the specific studies that are included in the evaluation. One must refer to the BenMAP manual to understand how the beta coefficients are derived, and as noted previously the manual (or prior assessments) do not include derivation of the coefficients for the specific studies included in the PA evaluation. Based on review of the BenMAP model, it is my understanding that the beta estimates are more in line with the first choice, that is "Beta estimates the percent change in the conditional expected *observed* value of the health impact *associated with* a unit change in the *observed* value of the pollution variable." Or as noted this would be most similar to a regression interpretation.

3. Similarly, is the definition of "concentration-response (C-R) relationships" in the PA and its Appendices (cf p. C-38) adequately clear and unambiguous to support simulation of well-defined causal effects of interventions that change pollution levels?

As with the beta coefficients, EPA does not provide a clear definition of the concentration-response (CR) functions other than that they are obtained from the epidemiological studies. The epidemiological studies that underly the CR functions are not intervention studies, but rather are observational studies that report associations between all-cause or cause-specific mortality and estimated ambient air concentrations.

4. Do the beta coefficients in the PA overcome these methodological objections to using relative risks, regression coefficients, and related measures of association to predict (or simulate) effects of interventions?

The PA is not explicit with regards to how the underlying epidemiological studies were chosen, and whether these studies reflect associations or true predictions that could be inferred from intervention studies. Furthermore, EPA does not provide a discussion of how important limitations associated with these epidemiological studies (e.g., exposure measurement error and confounding) impact the interpretation of the risk assessment results.

8. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately describe and discuss the extent (if any) to which their values reflect potential omitted confounders of the association between mortality risks and PM2.5 levels (e.g., lagged daily high and low temperatures and humidity in the weeks preceding mortality, if these contribute both to (possibly delayed) mortality and increased energy usage and PM2.5 pollution.

The discussion of uncertainty in the PA is limited. The PA only discusses potential confounding by co-pollutants qualitatively, noting the confounding could impact results in both directions and is of low-medium magnitude and further that most studies evaluated in the ISA found "relatively unchanged" effect estimates in co-pollutant models. EPA does not discuss the impact of other potential confounders, including meteorological parameters. This is an important source of uncertainty that needs to be included in the PA.

There are several examples in the literature that suggest confounding by co-pollutants or other variables and to my knowledge this issue is far from being resolved as suggested by the discussions in the PA. For example, Burnett *et al*. (1997) evaluated confounding by gaseous co-pollutants in a study of hospital

admissions (HA) and cardiovascular and respiratory disease in Toronto, Ontario from 1992 to 1994. The authors observed that in multi-pollutant models, PM contributions were reduced and not statistically significant. The authors concluded that "PM mass and chemistry could not be identified as an independent risk factor for the exacerbations of cardiorespiratory diseases in this study beyond those attributable to climate and gaseous air pollution." Similarly, in an independent mortality study of short-term NO_2 exposure and non-accidental mortality, $PM_{2.5}$ was associated with mortality in single-pollutant models, but not when it was adjusted for NO_2 (Burnett *et al*., 2004). Lastly, Klemm *et al*. (2004) demonstrated that all-cause mortality estimates associated with short-term $PM_{2.5}$ exposures in counties in Georgia were highly sensitive to temporal smoothing and confounding by co-pollutants. The authors concluded that "none of these findings supports a causal interpretation of the association between $PM_{2.5}$ and all-disease elderly mortality." The authors suggested that $PM_{2.5}$ was acting as a surrogate for other environmental agents.

Burnett, RT; Cakmak, S; Brook, JR; Krewski, D. 1997. "The role of particulate size and chemistry in the association between summertime ambient air pollution and hospitalization for cardiorespiratory diseases." *Environ. Health Perspect.* 105(6):614-620

Burnett, RT; Stieb, D; Brook, JR; Cakmak, S; Dales, R; Raizenne, M; Vincent, R; Dann, T. 2004. "Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities." *Arch. Environ. Health* 59(5):228-236.

Klemm, RJ; Lipfert, FW; Wyzga, C; Gust, C. 2004. "Daily mortality and air pollution in Atlanta: Two years of data from ARIES." *Inhal. Toxicol.* 16(Suppl. 1):131-141

It is more complicated to control for confounding in long-term studies than in short-term (time-series) studies because inferences from long-term studies are based on differences in pollution levels between cities, as opposed to day-to-day differences in pollution levels in a single city. Thus, factors such as socioeconomic (such as poverty levels) or general lifestyle factors (such as smoking or obesity rates) that vary from city to city could be a potential confounders. Controlling for these factors is difficult and residual confounding is likely almost always present because of lack of data, even if surrogates or some form of the variable can be included in the statistical models. For example, analyses of the ACS data (e.g., Krewski *et al*. 2009), featured analyses that addressed potential biases and confounding associated with various sociodemographic ecological factors and spatial model specifications, which included controlling for individual-level covariates. The information for the individual, however, were taken from the original 1982 enrollment questionnaire, and no follow-up was conducted. In addition, neighborhood-level socioeconomic covariates (*e.g.*, education, poverty levels, and unemployment) were based on census data that also likely varies over time. EPA continues to heavily rely on the ACS cohort analyses, although issues with confounding have not been resolved.

Krewski, D; Jerrett, M; Burnett, RT; Ma, R; Hughes, E; Shi, Y; Turner, MC; Pope, AC III; Thurston, G; Calle, EE; Thun, MJ. 2009. "Extended Follow-up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality." HEI Research Report 140. Health Effects Institute, Cambridge, MA

9. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address the extent (if any) to which their values reflect residual confounding of the association between mortality risks and PM2.5 levels (e.g., by daily high

and low temperatures and humidity in the weeks preceding mortality in models that only address seasonal, annual, or averaged temperatures)?

As with other potential confounders, EPA does not address the impact of potential residual confounding in the epidemiological studies. This should also be discussed in the PA. Importantly, it should be part of a quality and risk of bias evaluation in the PM ISA.

10. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address model uncertainty (e.g., the possibility that the linear no-threshold model specification is incorrect, e.g., because sufficiently low exposure concentrations do not cause pulmonary inflammation and adverse health effects that occur at higher concentrations)?

Model uncertainty remains one of the largest sources of uncertainty in observational air pollution studies. For example, adjustments for temporal trends (with respect to other population behaviors or episodic diseases such as viral infections), control for climate and weather effects, and confounding from co-pollutants have been addressed in a comprehensive re-analysis of seminal time-series studies (*e.g.,* HSC study and NMMAPS), which include extensive sensitivity analyses. The results from these analyses showed that methods used for controlling temporal trends and weather can yield large variations in effect estimates (HEI, 2003). Both HEI and Moolgavkar (2005) stressed that there is no objective statistical test to determine the best method to control for these factors. Analyses of $PM_{2.5}$ effects have been shown to be particularly sensitive to selection of degrees of freedom used in smoothing functions to control for temporal confounding in time-series studies (Dominici *et al*., 2007; Ostro *et al*., 2008).

Dominici, F; Peng, RD; Zeger, SL; White, RH; Samet, JM. 2007. "Particulate air pollution and mortality in the United States: Did the risks change from 1987 to 2000?" *Am. J. Epidemiol.* 166(8):880-888.

Health Effects Institute (HEI). 2003. "Revised Analyses of Time-Series Studies of Air Pollution and Health: Revised Analyses of the National Morbidity, Mortality and Air Pollution Study, Part II. Revised Analyses of Selected Time-Series Studies." Health Effects Institute, Cambridge, MA, May.

Moolgavkar, SH. 2005. "A review and critique of the EPA's rationale for a fine particle standard." *Regul. Toxicol. Pharmacol.* 42:123-144.

Ostro, B; Broadwin, R; Green, S; Wen-Ying, F; Lipsett, M. 2008. "Fine particulate air pollution and mortality in nine California Counties: Results from CALFINE." *Environ. Health Perspect*. 114(1):29-33.

11. Does the PA's discussion of beta values, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address effects on the estimated values of the beta values of exposure uncertainties and estimation errors (e.g., the possibility that individual exposure concentrations among people with adverse health responses tend to be higher than those among people who did not respond, even when they have the same estimated exposure values)?

Exposure measurement error is one of the largest sources of bias in long- and short-term ecological epidemiological studies. Most studies rely on central-site monitors as surrogates for personal PM exposure

or newer hybrid models that rely on monitored and modeled concentration, which do not necessarily accurately capture population mobility, or capture uneven distribution of PM attributable to local sources (monitoring sites may represent a nearby source and not human exposures a small distance away), pollution patterns that can be affected by terrain features and weather, or daily variations in PM concentrations or composition that may differ from variations experienced by individuals. Sarnat *et al*. (2009) found that personal exposures to sulfate (a major component of $PM_{2.5}$ in certain parts of the country) average over time; varied by individual, city, and season; and that this variability can lead to C-R functions that do not represent the true relationship between exposure and outcome.

> Sarnat, JA; Brown, KW; Bartell, SM; Sarnat, SE; Wheeler, AJ; Suh, HH; Koutrakis, P. 2009. "The relationship between averaged sulfate exposures and concentrations: results from exposure assessment panel studies in four U.S. cities." *Environ. Sci. Technol.* 43(13):5028-5034.

12. Does the PA adequately assess the suitability of the designs of the studies used to estimate beta values for purposes of valid causal inference and simulation? Does the PA's discussion of uncertainty and sensitivity analyses adequately address the internal validity and external validity (generalizability) of the estimated beta values used to simulate the causal impacts on public health risks of changing PM2.5 levels?

As noted previously, in the ISA, EPA did not include a systematic evaluation of study quality and risk of bias of individual epidemiology studies, included the studies EPA relies on in the PA for the risk assessment and does not discuss issues of internal or external validity associated with these studies. This is an aspect that is missing from the overall evaluation of this evidence.

13. Does the PA's discussion of beta values adequately address attribution of risk in the presence of joint causes? For example, if a unit change in PM2.5 levels has different expected effects on mortality risk for a person below the poverty line and during extremely hot or cold weather than it would for an initially similar (exchangeable) person with higher income and no exposure to extreme temperatures, then how much of the statistical "effect" of PM2.5 on mortality risk, as reflected in the beta values in Table C-1, should be attributed to income and weather variables? Conversely, how well does the discussion in the PA make clear how much of the estimated beta value for PM2.5 is actually contributed by other variables (such as temperature extremes and poverty) that would not necessarily be changed by an intervention that reduces PM.5 levels?

Although sensitivity analyses have been conducted to evaluate whether differences in socioeconomic status or extreme temperature explain the heterogeneity in the mortality risk estimates across regions and seasons, the issue of observed heterogeneity across regions and seasons has not been resolved. Also, using National estimates from multi-city studies tends to mask this heterogeneity and is likely inappropriate for use when evaluating effects in specific cities. For example, in an analysis conducted by Krewski *et al.* (2009), the authors evaluated mortality in two cities, Los Angeles and New York City using estimates of PM2.5 exposure based on land-use regression (LUR) models and/or kriging methods. Despite the use of similar methods to estimate $PM_{2.5}$ exposures in the two cities, the city-specific mortality risks differed. The LA mortality risks were only statistically significant for all-cause and ischemic heart disease mortality for estimates were adjusted for 44 individual-level covariates, but not for COPD or lung cancer. None of the mortality risk estimates were statistically significant in NYC. These results indicate that mortality risks are not consistently positive or significant across cities and the CR functions (or betas) would differ and even be null if city-specific estimates were used in the BenMAP analysis rather than National estimates. In fact, based on the intra-urban study results, the

authors concluded that "comparing the mortality risk estimates obtained from the Nationwide Analysis with those from the Intra-Urban Analyses indicates that the Nationwide risk estimates cannot be directly applied to all urban areas within the United States and that mortality risk estimates can vary appreciably among large urban areas with different characteristics" (Krewski *et al.*, 2009)

Overall, the beta values used in the BenMAP analyses are from studies that likely did not effectively control for variables that potentially could confound or entirely explain the relationship between PM2.5 and mortality. This is an aspect of uncertainty in the risk assessment that needs to be further discussed in the PA with regards to interpreting the results.

Furthermore, going back to the ISA, EPA needs to include a discussion of alternative hypotheses that could explain the observed associations in observational epidemiological studies, particularly given the very small effect estimates. For example, in all-cause mortality, a large portion of the mortality is attributed to cardiovascular mortality. Cardiovascular mortality has a large number of risk factors and it is well-known that activity and stress can trigger heart attacks. It is plausible that stress can confound relationships between day to day air pollution and mortality (i.e., related to driving and commuting) particularly in urban areas.

Krewski, D; Jerrett, M; Burnett, RT; Ma, R; Hughes, E; Shi, Y; Turner, MC; Pope, AC III; Thurston, G; Calle, EE; Thun, MJ. 2009. "Extended Follow-up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality." HEI Research Report 140. Health Effects Institute, Cambridge, MA

14. Overall, does the PA and its underlying documents (e.g., the BenMAP-CE documentation) make a convincing technical case that its simulated health impacts of reductions in PM2.5 are trustworthy and usefully accurate? How confident can policy analysts and decision makers be in the predictive validity of the simulated results?

The PA would benefit from a more detailed discussion of the uncertainties in the risk estimates that are presented, and in particular it needs to be more explicit regarding the assumption of causality. That is, that if the relationship is not causal then the risk estimates are not valid. Also presenting risk estimates down to a zero level of air pollution is misleading and unrealistic, given that there is always some level of PM2.5 that can't realistically be controlled. If the risk assessment is included in the PA, the focus should be on the comparison between current or just meeting the NAAQS with alternative NAAQS and should include a statistical analysis to assess whether *differences* in risk estimates are statistically different between different scenarios given full consideration of the uncertainties (not limited to the statistical uncertainty in the beta coefficient).

15. Are there other statistical or methodological issues that you would like to comment on that you believe might help the CASAC to assess the validity and soundness of the PA and its simulations for effects on health risks of changing PM2.5 levels, or that might help to improve the technical and scientific quality of the final PA?

As noted previously, the uncertainties in the BenMAP analyses are limited and should be expanded to include other sources of uncertainty, including but not limited to the modeling uncertainty in the ambient air concentrations for the scenarios of just meeting the NAAQS or alternative NAAQS (see for example, Table 1 from Dr. Lange's questions).

16. How can techniques of formal causal modeling and analysis best be applied to improve the clarity of definitions and communication and scientific soundness of simulations, inferences, causal interpretations, generalizations, and policy-relevant conclusions in the PA? Please comment on whether any aspects of the following (or other) causal model formalisms can substantially improve the clarity and scientific soundness of the analyses and simulations in the PA: causal graph and DAG methods, conditional independence tests, intervention and interrupted time series analyses, other quasi-experimental methods, Wiener-Granger causality and transfer entropy, causal dynamic Bayesian networks (DBNs), other information-theoretic and graph methods, Simon-Iwasaki causal ordering , non-parametric structural equations models, mediation analysis.

There are many aspects of the EPA causal framework that could be improved to provide greater transparency and scientific soundness to the NAAQS process. For example, the improvements would start as part of the development of systematic review protocol that follows modern systematic review methodology, including a solid approach for assessing study quality and risk of bias and ranking a weighing studies based on study quality and risk of bias findings. In addition, EPA's integration of the evidence could be improved such that no one line of scientific evidence (e.g., epidemiological evidence) is used exclusively to draw conclusions, especially when other lines of evidence are not consistent or coherent with the selected line of evidence.

Also, with regards to specific lines of evidence, in this case epidemiological evidence, EPA needs to provide more clarity regarding the inclusion and exclusion of studies. As was noted by CASAC in it review of the draft PM ISA, EPA excluded many studies that were informative regarding causal associations between PM and various health effects, including many so called "accountability" studies. Similarly, EPA tends to exclude occupational studies, even though these studies can also provide important information regarding health effects at high levels of PM exposure and whether this evidence is consistent and coherent with observational population-based studies.

I cannot comment specifically on formal causal modeling and analyses techniques as these are outside my expertise.

Questions from Dr. Sabine Lange

1. Is it appropriate to compare daily PM2.5 concentrations to the annual average?

As presented in Appendix B of the draft PA, it appears that EPA calculates both annual average pseudo-design values and 24-hr maximum pseudo-design values. However, as presented in Figure 3-9, EPA appears to be presenting graphically only the annual average pseudo-design values for both short-term and long-term studies (although this is not clear from the text or from the Figure caption). I agree that it would be most appropriate for EPA to evaluate long-term studies against annual average design values, and short-term studies against the the 24-hr design values. It is unclear why EPA is presenting only the annual average design values when discussing short-term studies.

2. Is it informative to derive annual average pseudo-design values for study areas in short-term studies (that look at effects of day-to-day PM2.5 concentration changes), in order to determine whether these study areas attained the current annual standard? Although the EPA can technically determine if daily changes in PM2.5 concentration increased health effects in an area meeting the annual standard, does this really inform the health protectiveness of the annual standard? It seems that whether an area showed a positive effect or not could be completely independent of the annual standard and instead dependent on how much the PM2.5 concentrations changed from day-to-day.

In the PA, EPA acknowledges that the estimated ambient concentrations used in the epidemiological studies that it relies on to evaluate the adequacy of the PM2.5 NAAQS are not the same as the design values that are used to determine compliance with the NAAQS, which is true. To address this, EPA therefore decided to calculate pseudo-design values to compare with the reported ambient concentrations in the epidemiological studies and determine whether the study areas in the epidemiological studies would have met or violated the current NAAQS. Unfortunately, this does not address the more fundamental question of how the estimated ambient concentrations in the epidemiological studies actually reflect individual exposures to PM, and how the likely exposure measurement error impacts the reported association between PM and various health effects or the shape of the concentration-response function. Instead, EPA appears to use this analysis to conclude that most of the selected studies (18 of 29) – covering both short-term and long-term PM2.5 exposures - that observed positive associations between PM2.5 and mortality and morbidity endpoints had some portion of the population (25-75%) that lived in an area where the air quality met current NAAQS, and for the other 11 studies a majority of the population (> 75%) lived in areas that did not meet the NAAQS. Looking at Figure 3-9, it appears that the few studies had pseudo-design values that were mostly below the NAAQS, the large majority included some values that exceeded the NAAQS, making it difficult to interpret these results in terms of assessing the adequacy of the NAAQS. As noted by EPA, for many of the multi-city studies some locations would likely have met the NAAQS and others would not. It is unclear how useful this analysis is for determining the health protectiveness of the current NAAQS, especially when relying solely on this epidemiological data.

3. In contrast to short-term studies that investigate the effects of day-to-day changes in PM2.5 concentrations within a certain geographic area, long-term cohort studies often look at the association between annual average PM2.5 concentrations and time-to-event data (such as the time from cohort entry to death) over long periods of time. For these studies, it is not uncommon for all study subjects in a single geographic area to have the same (or very similar) exposure assignments (e.g. Jerrett et al., 2017; Thurston et al., 2016), in which case the study is assessing the effects of PM2.5 between geographic areas, instead of within geographic areas. In this case, is the pseudo-design value in a

single geographic area particularly informative, when the association between PM2.5 and the health effect is driven by the differences between study areas?

As noted above, it is unclear how comparing the pseudo design values for the locations in underlying long-term epidemiology studies (some which are below and others that are above the current NAAQS) informs the adequacy of the current NAAQS. As noted by Dr. Lange, this is further complicated by the fact that the underlying epidemiology study is not assigning exposures using the same criteria as EPA. As noted by EPA, additional uncertainties include the number of monitors that are included in the calculation of the design values that may not reflect the same monitors used in the underlying epidemiology study. EPA should clarify how this analysis is informative for addressing policy questions, given the uncertainties and the clear disconnect between how exposures are estimated in the epidemiology studies and how EPA determines compliance with the NAAQS.

4. Is there a quantitative uncertainty analysis method that the EPA could use for this risk assessment that captures more of the uncertainty and variability of the risk estimates (such as those described in Table 1), in order to better inform CASAC and the EPA Administrator about the impact of these uncertainties?

BenMAP uses Monte Carlo analyses from which 95% confidence intervals are derived which incorporates only the statistical uncertainty (the standard error of the beta coefficients) in the risk estimates. EPA could expand this Monte Carlo analysis to include other sources of uncertainty – such as the uncertainty in the variables described in Table 1.

5. Could different magnitudes of error amongst different variables in regression analyses be masking the effect of a speciated constituent of PM2.5?

I am not sure if I understand the question. The differences in PM composition across regions has been hypothesized to account for the large heterogeneity in effect estimates across regions. However, I do not know of any study that has identified a specific PM constituent that is associated with the observed overall PM effects. It seems reasonable that the exposure measurement error would vary significantly for individual constituents, but it is unclear whether using PM mass masks any effect of individual constituents. Evaluations of individual constituents (such as sulfate) both in experimental and epidemiology studies do not necessarily support the adverse effects observed in studies that evaluate PM2.5 mass.

6. What happens when multiple potential explanatory factors are included in a single variable in an already-complex multiple regression system? Presumably each PM2.5 component has a different C-R relationship with the health effect (even if that relationship is zero), and each is a somewhat better or worse surrogate for the relationship between actual exposure vs measured exposure. What kind of an impact would this inclusion of multiple potential explanatory factors into one variable have on the final C-R function, and how accurate would that C-R function be?

This is an area of uncertainty that needs more attention and study. As noted above studies that have tried to evaluate specific constituents of PM have generally reported very inconsistent evidence and there is not clear single component that appears to explain the associations that are observed for PM mass. This is further complicated by the fact that studies have reported statistically significant associations not only between PM and mortality, but also between other criteria air pollutants (nitrogen dioxide, carbon monoxide, sulfur dioxide, and ozone) and mortality (e.g., Stieb *et al*., 2002), yet all of these air

pollutants are rarely included in recent epidemiology studies as potential confounders. And this does not include the myriad of other air pollutant (e.g., benzene, formaldehyde) that correlate with the criteria air pollutants. There is more inconsistency in the literature regarding confounding effects of co-pollutants than EPA generally recognizes in discussing this issue and this should be more fully addressed.

> Stieb, DM; Judek, S; Burnett, RT. 2002. "Meta-analysis of time-series studies of air pollution and mortality: Effects of gases and particles and the influence of cause of death, age, and season." *J. Air Waste Manage. Assoc.* 52:470-484.

Questions from Dr. Steven Packham

Are there areas (e.g., specific aspects of biological causation, pulmonary toxicology, or causality in epidemiology) in which additional knowledge is required to appraise the adequacy and basis of existing, new, or revised PM NAAQS to protect public health with an adequate margin of safety?

The NAAQS process in general could benefit from a more systematic process of evaluating individual studies assessing study quality and weighing and ranking studies. This would allow for a more balanced consideration of the evidence and a more transparent and objective analysis. Similarly, better integration of the evidence is needed in order to assess consistency and coherence in the data from different lines of evidence (epidemiological and experimental). The evidence is currently presented in the PA, heavily weighed in favor of the epidemiological evidence. These studies are not sufficient for assessing causality because of large inherent limitations including exposure measurement errors, uncontrolled confounding, among other issues. Information derived from controlled exposure studies and animal studies is apparently dismissed in the PA. For example, EPA notes that animal toxicology studies assess effects to PM2.5 concentrations ”well-above the concentrations likely to be allowed by the current PM2.5 standards (pg. 3-49 in the PA)” and therefore based on this and challenges with from extrapolating from effects in animals to humans, EPA concludes that “animal toxicology studies are of limited utility in informing conclusions on the public health protection provided by the current or alternative primary PM2.5 standards,” similarly, the for controlled human exposure studies EPA concluded:

“the PM2.5 exposure concentrations evaluated in most of these studies are well-above the ambient concentrations typically measured in locations meeting the current primary standards. Therefore, controlled human exposure studies provide limited insight into the occurrence of cardiovascular effects following PM2.5 exposures likely to occur in the ambient air in areas meeting the current primary PM2.5 standards and are of limited utility in informing conclusions on the public health protection provided by the current standard.”

I don’t agree with the EPA conclusions. Instead, I think there is utility in evaluating all lines of evidence to determine if these lines of evidence provide consistent and coherent evidence of an effect at concentrations lower than the current NAAQS. Both animal studies and controlled exposure studies have been used in numerous evaluations (by EPA and others) for assessing levels of no or low adversity.

Can you suggest additional specific scientific disciplines and areas of biomedical informatics and research (e.g., systems biology methods for clarifying biological causal pathways, mechanisms, modes of action, quantitative causal dose-response relationships) that should be included in future reviews of other criteria pollutants?

If EPA deems that the animal and controlled exposure studies are inadequate for informing the adequacy of the level of the NAAQS, it stands to reason that more studies of this kind at relevant ambient exposure levels are needed. It would also be important to conduct animal studies that evaluate not only markers of early effects (such as inflammatory reactions), but the complete chain of events that might result in overt disease or even mortality at relevant ambient exposures (i.e., equivalent levels below the current NAAQS).

Can you describe the research efforts necessary to provide the required information? To what extent has the needed research already been done, or started? For example, are there crucial experiments or research initiatives that could clarify the shape of the PM2.5-chronic inflammation causal dose-response relationships at relevant exposure concentrations? Are there specific data analyses (e.g., testing for confounding by weather variables over more days prior to mortality) that could clarify the causal interpretation of epidemiological associations relied on in the draft PA to simulate effects of interventions?

There is a large extensive body of literature that I believe already can be used if integrated properly and evaluated objectively to assess consistency and coherence across scientific lines of evidence. Solely relying on the epidemiological findings is not sufficient to establish causality or to determine a level of exposure with an adequate margin of safety (especially because as EPA notes there is no discernible threshold level that can be identified in epidemiology studies). The epidemiology studies are hindered because both the exposure and outcome are very common (we are all exposed to PM and we all die) and this makes the analysis challenging in the face of many variables that can explain the observed associations. The issue is further complicated by the fact that PM composition varies and there is no single or group of constituents that has been found to be the 'causative' component.

EPA identifies areas of future research and I agree that these areas of research could potentially aid in assessing future evaluations. These include (as noted above) toxicological studies that evaluate the full chain of events that would results in disease, long-term animal studies at low PM levels, and longer controlled human exposure studies at low PM levels. These studies could include PM of varying composition to help elucidate whether there are differential effects depending on composition.

Can you suggest additional areas of scientific literature review on species and individual human organism's capacities of adaptation to inhaled environmental stressors that might help establish margins of safety when exposed to ambient levels of air pollution?

One scientific line of evidence that I can think of that is missing in the evaluation of PM effects is the occupational literature. This literature has been included in past assessments to a limited degree (for example workers exposed to diesel particulate matter). I think inclusion of these studies can provide an important perspective and can also help identify levels of exposure that have been shown to be hazardous in these population groups (compared to recent observational population studies).

Could you provide additional information on the relative contributions to air pollution concentrations and resulting health effects of natural and anthropogenic activity? For example, is either one alone, or are both natural and anthropogenic activities together, sufficient to cause the magnitudes of adverse health effects attributed to PM2.5 in the Draft PA?

This is an interesting question, and I am not sure what the answer is. EPA does not distinguish between concentrations of PM from natural and anthropogenic sources and in fact assessed risks down to a zero

(non-existent) PM concentration. EPA does discuss "background" level of PM in the PA, but these background levels are not considered in the interpretation of the scientific evidence or in the risk assessment. In addition, PM from different sources will vary in composition and to date there is no specific PM component of group of components that have been identified as the causative agents that could account for the observed associations between PM and mortality.

Could you provide additional information on any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of such national ambient air quality standards?

This is outside my area of expertise, however, along with the NAAQS process, EPA also conducts a Regulatory Impact Assessment (also using BenMAP) that not only quantifies the health impacts, but also calculated the value associated with the health impacts (or avoided health impacts from for example lowering the NAAQS). These health impacts are driven by mortality and the value of a "statistical life" and far exceed estimated costs of complying with the NAAQS regulations.

Questions from Dr. Mark Frampton

1. Some CASAC members are of the opinion that the scientific evidence base is insufficient to support a causal relationship between PM exposure and mortality. Is there evidence that would support a reconsideration of the current and long-held views of this causal relationship as expressed in the PA and ISA?

I agree that the observational epidemiological evidence alone is insufficient to establish a clear causal relationship between PM2.5 and health effects at levels below the current NAAQS. The study designs cannot overcome the challenges of exposure measurement errors and potential confounding by a myriad of variables – particularly when evaluating a very common exposure and a very common outcome (i.e., mortality). All that the epidemiological literature can tell us is that there is a relatively small statistical association between estimated ambient PM2.5 concentrations and various health outcomes. These associations vary in precision and magnitude on both temporal and spatial scales (i.e., by city or region and by season), and yet this heterogeneity does not appear to be related with PM composition, or even with relative PM concentrations (i.e., large health effects are often observed in areas with lower PM concentrations compared to higher PM concentrations). The observational epidemiology literature also is not consistent with the occupational epidemiological literature. For example, in occupations associated with high PM exposures (carbon black, coal, talc, etc.) workers may develop respiratory illnesses, but high rates or cardiovascular diseases or all-cause mortality are not observed in these populations, despite being exposed to levels several fold higher than ambient concentrations. Similarly, controlled exposure studies and animal studies indicate inconsistent results at high PM exposures.

2. In the long-term epidemiological studies of PM mortality, heterogeneity between and within cities has been cited as a source of uncertainty in drawing conclusions about causality. Please opine on the level of uncertainty that is represented by this heterogeneity, and the impact if any, on the conclusions in the PA.

As noted above and copied here:

Although sensitivity analyses have been conducted to evaluate whether differences in socioeconomic status or extreme temperature explain the heterogeneity in the mortality risk estimates across regions and seasons, the issue of observed heterogeneity across regions and seasons has not been resolved. Also, using National estimates from multi-city studies tends to mask this heterogeneity and is likely inappropriate for use when evaluating effects in specific cities. For example, in an analysis conducted by Krewski *et al.* (2009), the authors evaluated mortality in two cities, Los Angeles and New York City using estimates of PM2.5 exposure based on land-use regression (LUR) models and/or kriging methods. Despite the use of similar methods to estimate $PM_{2.5}$ exposures in the two cities, the city-specific mortality risks differed. The LA mortality risks were only statistically significant for all-cause and ischemic heart disease mortality for estimates were adjusted for 44 individual-level covariates, but not for COPD or lung cancer. None of the mortality risk estimates were statistically significant in NYC. These results indicate that mortality risks are not consistently positive or significant across cities and the CR functions (or betas) would differ and even be null if city-specific estimates were used in the BenMAP analysis rather than National estimates. In fact, based on the intra-urban study results, the authors concluded that "comparing the mortality risk estimates obtained from the Nationwide Analysis with those from the Intra-Urban Analyses indicates that the Nationwide risk estimates cannot be directly applied to all urban areas within the United States and that mortality risk estimates can vary appreciably among large urban areas with different characteristics" (Krewski *et al.*, 2009)

3. Please opine on the adequacy of the causality analysis framework currently used by the EPA, and whether and how the concepts espoused by Dr. Cox should, or should not, be incorporated into the NAAQS causality framework. Also please comment on the implications, of any changes in the causality framework that you would recommend, for the analyses and conclusions in this current PA.

As noted above and copied here (with some additions):

The NAAQS process in general could benefit from a more systematic process of evaluating individual studies for assessing study quality and weighing and ranking studies and this is currently lacking the NAAQS framework. This would allow for a more balanced consideration of the evidence and a more transparent and objective analysis (i.e. instead of relying so heavily on one line of evidence – such as the epidemiology evidence). Similarly, better integration of the evidence is needed in order to assess consistency and coherence in the data from different lines of evidence (epidemiological and experimental) and this is also lacking as applied under the current framework. The evidence is currently presented in the PA, heavily weighed in favor of the epidemiological evidence. These studies are not sufficient for assessing causality because of large inherent limitations including exposure measurement errors, uncontrolled confounding, among other issues. Information derived from controlled exposure studies and animal studies is apparently dismissed in the PA, even though this type of evidence is often used to derive health-based limits for numerous chemicals.

I cannot specifically comment on the causal methodologies that Dr. Cox has recommended as this is outside my area of expertise. However, a more balanced assessment of the scientific evidence would likely result in changes to the overall conclusions in the PA, or at least provide more context that would allow for a more critical interpretation of the data.

4. "There is inadequate evidence for the 'likely to be causal' conclusion for long-term PM2.5 exposure and cancer." This is based on epidemiological studies that do not appear to adequately differentiate incident cancer and cancer-related mortality because the exposure time frames for most of these studies

are insufficient to draw conclusions about incident cancer. Do you agree with the CASAC's findings in this matter? Please discuss the evidence (or lack thereof) that supports your opinion.

In the previous EPA PM evaluation, EPA concluded that evidence was suggestive but not enough to conclude that it is "likely to be causal" and this is largely because of the lack of consistent finding in the epidemiological literature and the lack of support of *in vivo* mutagenicity or carcinogenicity in experimental animals. I agree that the literature still does not support a "likely to be causal" conclusion and that many studies do not provide an adequate lag time to establish temporality of effect (that is sufficient time that has elapsed between exposure and effect). As noted by Dr. Lange in her CASAC comments this includes the seminal studies of the H6C and CCHS cohorts.

5. Please comment on the appropriateness and completeness of the approaches used in this PA to assess the risks of exposure, and the assessments of risk reduction of alternate standards, for PM2.5 and PM10 (sections 3 and 4, respectively).

The BenMAP model relies solely on the selected epidemiological studies and these studies are not sufficient to confirm that there is a causal association between PM2.5 at current levels of exposure and mortality because of the inherent limitations of these studies (most importantly confounding and exposure measurement issues). EPA also estimates effects down to an unrealistic zero level of PM, which is misleading. Another important issue is that EPA applies National-level effect estimates to specific cities, even though studies have observed substantial heterogeneity in effect estimates across cities. Lastly, EPA only presents a limited uncertainty analysis that incorporates only the statistical uncertainty in the effect estimate derived from the epidemiological study. Other important sources of uncertainty are not quantified.

Questions from Dr. James Boylan

Chapter 2 – PM Air Quality

Is the discussion on sources of emissions accurate and complete? If not, what additional information needs to be included?

In general, the discussion is accurate, although somewhat limited and basic. Additional information that should be included in this section is information on how the PM emissions vary by region and also by season. In addition, there should be some discussion about the uncertainties in estimating the sources and emissions of PM.

Is the discussion on ambient monitoring accurate and complete? If not, what additional information needs to be included?

In general, this section is also accurate, but also somewhat limited. It would be useful to have information regarding the sampling schedule and how this has changed (or not) over time. For example, PM measurements taken every 3 or 6 days. It would also be useful to have information on how the ambient monitors are used for determining compliance with the NAAQS. I did not see this information in this section.

Is the discussion on ambient measurement correlations and trends accurate and complete? If not, what additional correlations and trends need to be included?

I found this section to be useful and adequate.

To what extent are biases associated with PM10, PM2.5, and ultrafine measurements discussed? How would differing PM2.5 biases associated with FRM vs. FEM continuous measurements (e.g., FEMs typically show higher PM2.5 concentrations compared to FRMs) impact the evidence-based and risk-based PM2.5 assessments in Chapter 3?

This is an area of uncertainty that is not included in either the epidemiological studies or in the risk assessment, although I would imagine that other sources of uncertainty would likely be much higher than the biases related to measurement errors. For example, exposure measurement errors are likely to be higher in epidemiological studies, and modeling uncertainty and potential biases are likely to be higher in the risk assessment.

Is the discussion on hybrid modeling approaches accurate and complete? If not, what additional information needs to be included?

Overall, I think this section was adequate. However, one area that I think is critical for evaluating exposure measurement error in the epidemiological studies is how ambient concentrations (whether measured or modeled) relate to personal exposures in individuals that spend time in different microenvironments. This is additional information that could be included in this section.

Is the discussion on performance methods for evaluating hybrid modeling methods accurate and complete? If not, what additional information needs to be included?

I am not familiar enough with these models to opine on the adequacy of the performance evaluations.

Is the discussion on background concentrations accurate and complete? If not, what additional information needs to be included?

I was surprised by the low background levels of PM reported in this section. While I have not reviewed the literature on this issue, as with the question above, one important questions is how ambient air pollution levels relate to personal exposures. One can imagine that any individual person may have very different background levels of exposure depending on their personal activities, workplace exposures, and proximity to different natural sources of PM (e.g., wildfires).

Chapter 3 – Review of the Primary PM2.5 Standards

Is the evidence-based analysis presented in Chapter 3 scientifically sound?

As noted above and copied here:

The NAAQS process in general could benefit from a more systematic process of evaluating individual studies for assessing study quality and weighing and ranking studies and this is currently lacking the NAAQS framework. This would allow for a more balanced consideration of the evidence and a more transparent and objective analysis (i.e. instead of relying so heavily on one line of evidence – such as the epidemiology evidence). Similarly, better integration of the evidence is needed in order to assess consistency and coherence in the data from different lines of evidence (epidemiological and experimental) and this is also lacking as applied under the current framework. The evidence is currently presented in the PA, heavily weighed in favor of the epidemiological evidence. These studies are not sufficient for assessing causality because of large inherent limitations including exposure measurement errors, uncontrolled confounding, among other issues. Information derived from controlled exposure studies and animal studies is apparently dismissed in the PA, even though this type of evidence is often used to derive health-based limits for numerous chemicals.

Is the risk-based analysis presented in Chapter 3 and Appendix C scientifically sound?

As noted above and copied here:

> The BenMAP model relies solely on the selected epidemiological studies and these studies are not sufficient to confirm that there is a causal association between PM2.5 at current levels of exposure and mortality because of the inherent limitations of these studies (most importantly confounding and exposure measurement issues). EPA also estimates effects down to an unrealistic zero level of PM, which is misleading. Another important issue is that EPA applies National-level effect estimates to specific cities, even though studies have observed substantial heterogeneity in effect estimates across cities. Lastly, EPA only presents a limited uncertainty analysis that incorporates only the statistical uncertainty in the effect estimate derived from the epidemiological study. Other important sources of uncertainty are not quantified.

Is the discussion on following topics adequate and complete? If not, what additional information needs to be included?

o Study area selection,

I did not have any specific issues with the selection of study areas.

o Health outcomes (e.g., decision to focus on mortality and ignore cardiovascular and respiratory effects),

Previous evaluations have included morbidity outcomes and these are useful to provide context (i.e., higher ER visits and hospital admissions would be expected compared to mortality)

o Concentration-response functions,

While EPA provides some justification for the selection of studies, the justification does not include an evaluation of study quality or risk of bias. EPA should include more information on why it considered these studies to be of higher quality than other available studies.

o PM2.5 air quality scenarios evaluated,

It is unclear why EPA evaluates separately the scenarios for just meeting the NAAQS (vs. zero) and meeting alternative standards (vs. zero). EPA should consider evaluating the baseline vs. the control scenarios to estimate the difference directly.

o Model-based approach to adjusting air quality,

I am not familiar enough with the methodology to opine.

o Linear interpolation/extrapolation to additional annual standard levels, and

I am not familiar enough with the methodology to opine.

o Characterization of variability and uncertainty in the risk estimates.

The quantitative uncertainty characterization is extremely limited and should be expanded to include other sources of uncertainty (see for example Dr. Lange's questions).

Are the areas for additional research adequate and complete? If not, what additional areas need to be included?

Appendix C – Supplemental Information Related to the Human Health Risk Assessment

Is the air quality modeling approach to projecting PM2.5 concentrations to correspond to just meeting the NAAQS (AQS, CMAQ, Downscaler, SMAT-CE, project monitors to just meet NAAQS, project spatial fields to correspond to just meeting the NAAQS) scientifically sound? If not, what are your concerns and how should they be addressed?

I am not sufficiently familiar with the air quality modeling to opine on whether it is scientifically sound, however, the modeling uncertainty should be quantified and this uncertainty should be incorporated in the risk assessment.

Is the CMAQ model configuration and input files used in the air quality modeling appropriate for this application? If not, what updates are recommended?

I am not sufficiently familiar with the air quality modeling to opine on whether the input files are appropriate.

Are the CMAQ model performance metrics that were evaluated appropriate and adequate for this application? Are there any concerns with the model performance in any of the study areas used in the human health risk assessment? If so, how should these concerns be addressed in the health risk assessment?

I am not sufficiently familiar with the air quality modeling to opine on whether the model performance metrics are appropriate, but as noted above the modeling uncertainty should be incorporated in the calculation of the risk estimates.

Is the health risk modeling approach using BenMAP-CE appropriate for this application? If not, what are your concerns?

As noted above and pasted here:

The BenMAP model relies solely on the selected epidemiological studies and these studies are not sufficient to confirm that there is a causal association between PM2.5 at current levels of exposure and mortality because of the inherent limitations of these studies (most importantly confounding and exposure measurement issues). EPA also estimates effects down to an unrealistic zero level of PM, which is misleading. Another important issue is that EPA applies National-level effect estimates to specific cities, even though studies have observed substantial heterogeneity in effect estimates across cities. Lastly, EPA only presents a limited uncertainty analysis that incorporates only the statistical uncertainty in the effect estimate derived from the epidemiological study. Other important sources of uncertainty are not quantified.

Do the risk summary tables showing the impact of alternative PM2.5 standards and graphical plots showing the distribution of risk across ambient PM2.5 levels clearly and accurately summarize the results of the health risk analysis? If not, what additional information should be included?

Even though EPA only presents a limited uncertainty analysis, in the PA this uncertainty is presented as 95% confidence bounds around point estimates. When comparing results between attainment of current NAAQS and alternative NAAQS, however, EPA focuses on the point estimates and does not discuss the uncertainty bounds. EPA should provide statistical analyses to evaluate whether the calculated effect estimates differ significantly between meeting the current NAAQS and alternative NAAQS given the overlapping confidence bounds. The implications of this analysis on the risk conclusions should be included in the PA.