Preliminary Comments from Members of the

Independent Particulate Matter Review Panel

on

EPA's Policy Assessment for Review of the National Ambient Air Quality Standards for Particulate Matter (External Review Draft – September 2019)

Received as of 10-08-19

Dr. Douglas W. Dockery
October 7, 2019

SCQ-3.2 What are the panel's views on the relative weight that the draft Policy Assessment gives to the evidence-based (i.e. draft PA, section 3.2) and risk-based (i.e. draft PA, section 3.3) approaches in reaching conclusions and recommendations regarding current and alternative $PM_{2.5}$ standards?

Section 3.2 provides a well-structured and clearly presented synthesis of the evidence for the health effects of PM exposures. There is no evidence for a discernable population threshold. Two approaches are then used to attempt to draw out information relevant to recommending or evaluating current and alternative $PM_{2.5}$ standards.

In one approach, the $PM_{2.5}$ air quality distributions over which epidemiologic studies support health effect associations and the degree to which such distributions are likely to occur in areas meeting the current (or alternative) standards are evaluated. Studies are characterized based on the intuitive (but flawed) notion that the measures of association (exposure-response functions) are most precise at the mean of the exposure distribution. This misses the point that power is a function of the variance of exposure, not the mean. However, this approach does show evidence for $PM_{2.5}$ effects for studies with mean exposures below the current $PM_{2.5}$ standards.

In the second approach, $PM_{2.5}$ design values ("pseudo-design values") are calculated where possible for epidemiologic study sites. These calculations attempt to determine if these epidemiologic study areas would have met or violated the current or alternative standards during study periods. It is an interesting question whether the $PM_{2.5}$ exposure measures used in the epidemiologic studies (whether directly measured or estimated) would differ from the observed $PM_{2.5}$ from regulatory monitoring. Indeed, it is clear that regulatory monitoring by targeting compliance will produce values higher than monitoring or hybrid methods targeted on estimating population exposures. Ultimately, this approach also provides evidence for $PM_{2.5}$ effects in communities not violating the current standard, although the added value of this information is marginal.

Section 3.3 is a risk assessment that estimates population-level health risks associated with $PM_{2.5}$ air quality "requisite" to protect the public health, that is "just meeting" the current standards. Given evidence based conclusions of effect below the current standards, risks associated with $PM_{2.5}$ air quality adjusted to simulate "just meeting" alternative annual and 24-hour standards with lower levels are estimated. Although characterized as representave of the US population, this risk assessment is limited to 47 urban areas with Monitored $PM_{2.5}$ above of marginally below the current NAAQS. Mutiple urban areas affected by "special" circumstances such as wildfires, seasonal local wood smoke, and "uncertain" measurements ae excluded. A convoluted process estimated exposure reductions for each monitoring site to achieve targeted alternative based on a hybrid model of monitored and CMAQ model surfaces. The observed exposure response functions for a limited set of health outcomes ("causal" and "likely to be causal") are applied using BENMAP to estimate expected numbers of events. There is some quantitative, but largely qualitative assessment of uncertainty While this risk assessment provides some numbers, it does not substantially contribute to the recommendations regarding current and alternative $PM_{2.5}$ standards.

SCQ 3.3 What are the panel's views on the evidence-based approach, including:

a) The emphasis on health outcomes for which the draft ISA causality determinations are "causal" or "likely causal"?

The focus on the health outcomes which are "causal" or "likely causal" is entirely appropriate. (Note that the risk assessment approach only considers a subset of these health outcomes, see SCQ 3.4 a).

b) The identification of potential at-risk populations?

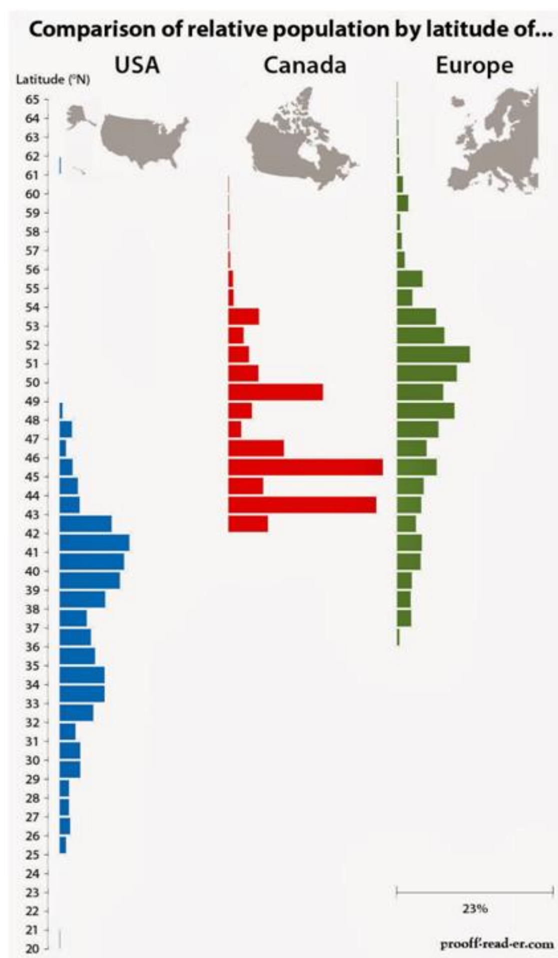
Section 3.2.2 (page 3-42) on "Potential At-Risk Populations" is remarkably succinct. It would be helpful to structure this discussion around the multiple pathways that people could be at risk because of exposure, susceptibility, ameliorating personal characteristics, and community context.

The evidence continues to be that the young, the old, and those with pre-existing chronic conditions have increased susceptibility. In addition, minority and economically disadvantaged populations have higher exposures and less ability to modify their exposure, to obtain appropriate health care, or to personal characteristics (e.g. nutrition) to ameliorate response. The conclusion is correct that *"the groups at risk of* $PM_{2.5}$*-related health effects represent a substantial portion of the total U.S. population*" (page 3-43, line 19).

c) Reliance on key multicity epidemiology studies conducted in the US and Canada for assessing the $PM_{2.5}$ ***levels associated with health effects?***

While there have been important and informative studies from Europe and Asia as well as from North America, the evidence from the multicity US and Canadian epidemiologic studies is adequate to assess the health effects of $PM_{2.5}$.

One might ask about the relevance of the Canadian studies, The population in Canada tend to be in the southern provinces which are further south than many US cities. While the 49th parallel is often thought of as the border between the US and Canada, the vast majority of Canadians (roughly 72%) live below it. Europe is further north, but has great overlap with US and Canada.



https://www.citymetric.com/horizons/we-think-canada-long-way-north-half-its-population-lives-south-milan-1194

d) Characterizing air quality in these key studies using two approaches: the overall mean and $25^{th}/75^{th}$ percentiles of the distribution and the "pseudo design value" reflecting a monitor with the highest levels in an area?

It is commendable to examine the distribution of the underlying $PM_{2.5}$ exposure data used in epidemiologic studies (page 3-51). Indeed there is useful information to be gained, particularly in considering how informative these studies are in the lower exposures ranges. However, examining these studies based on the mean exposure is not useful. Indeed, this approach is based on flawed understanding of the statistics.

The statement that epidemiologic studies provide the strongest support for reported health effect associations over the part of the distribution corresponding to the bulk of the underlying data (page 3-51, line 2-3) has same intuitive validity. However, extending that to say the associations are "strongest" at the mean is flawed. Figure 3-2 (page 3-52) from Lepeule et al (2012) is used to show that the confidence intervals are smallest at the center of the distribution of exposures (where there is the most data), and widest at the extremes. This is true, but this does not mean that the association is strongest (or alternatively has the smallest confidence interval) at the center of the distribution. The potted confidence intervals show the uncertainty around the expected hazard ratio at each exposure, and indeed these are larger where there is less data (or less exposure measures). In simple statistics, the error of the expected value is inversely proportional to the square root of the number of data points. Thus confidence intervals are wider where there is less data. However, the association is determined by the slope of the fitted line (note the expected value at any given point). In linear regression, the uncertainly (confidence interval) of the slope is inversely proportional to the square root of the number of data point times the variance of the exposure. Thus the important characteristics is not the mean of the exposure distribution but the variance or heterogeneity of the exposures. Studies with large differences in exposures are more precise than studies with little variation in exposure. A study with large numbers but no variation in exposure would produce a very precise estimate of the health indicator, but provide no information on the slope or association with exposure. Thus the parameter that should be examined in Figures 3-3, 3-4, 3-5, 3-6 and Table 3-3 is the variance or other index of heterogeneity (e.g. IQR) of exposure.

Likewise for Figure 3-7 and 3-8. Here it is positive that 20^{th} and 10^{th} percentiles are considered when available as well as mean. Indeed these percentiles would be a much more informative statistic to use in this risk assessment, but only a handful of these percentiles are available. Note that in these two figures, the arithmetic means could be compared to the short and long term studies. Indeed they show generally similar means. However, the variances and therefore the 10^{th} and 25^{th} percentiles should be very different, and cannot be directly compared. For the short-term studies variance is between daily $PM_{2.5}$ concentrations, and the number of data point is number of days. For long-term studies variance is between annual mean $PM_{2.5}$ concentrations, and the number of data points is the number of cities or spatial locations. Thus the short term studies will tend to have much larger variances than the long term studies.

There is a significant logical misinterpretation, or mislabeling, of the "pseudo design values". Throughout the PA there are statements such as "*50% of the study area populations lived in locations with pseudo-design values below these concentrations, or 50% of the health events occurred in such locations.*" This would appear to state that 50% of the population experiences such pseudo-design values, and equivalently 50% of the health events occur in these locations. Neither of these interpretations can be supported by the data.

e) The preference for continuing the use of an annual $PM_{2.5}$ ***standard as the principle means of providing public health protection against the bulk of the distribution of short- and long-term*** $PM_{2.5}$ ***exposures?***

The logic for this has not changed since the 2009 review. There is some additional discussion of this issue in the PA, which concludes there is no reason to modify this approach.

However, the increased frequency wildfires and acute exposures to anomalously high, short term episodes of $PM_{2.5}$ raises the importance of examining these effects in the evidence based analyses. (Note the risk assessment analyses explicitly exclude these events from consideration).

f) The draft PA conclusions on the extent to which the current scientific information strengthens or alters conclusions reached in the last review on the health effects of $PM_{2.5}$***?***

What are the changes in the evidence since the last review?

- Experimental studies (both controlled human and animal toxicology) providing evidence of causal pathways. Notably, some of these examine the same physiological and clinical indicators as in the epidemiologic studies. Exposures/doses in these experimental studies are higher than typically experienced as ambient exposures by populations in the community, requiring extrapolation, On the other hand, these exposures are now much closer (within an order of magnitude) of ambient.
- The hybrid methods for estimating exposures have allowed the epidemiology studies to examine not only populations living near a fixed monitoring station, but across larger regions or the entire country. Thus, studies now include complete samples of the population, not just those in urban areas where there were networks of air pollution monitors. The average levels in the US dropped over the last decade, so more communities have $PM_{2.5}$ levels below the current $PM_{2.5}$ NAAQS. Importantly now rural populations which were previously unmonitored are included. These rural populations tend to have lower exposures to $PM_{2.5}$, and therefore extend the range of observations to levels substantially below those included in the 2009 review. The national Canadian cohort studies have been particularly informative about effects at low $PM_{2.5}$ levels. However, the national cohort studies in the US have been able to examine associations restricting to communities with exposures below the current annual NAAQS.
- The hybrid methods have also improved the spatial resolution of the $PM_{2.5}$ estimates for epidemiologic analyses. These improved $PM_{2.5}$ exposure estimates have reduced exposure misclassification, effectively increased the sample size, and provided stronger, more precise associations.

All of these strengthen the evidence for health effects of $PM_{2.5}$ exposures.

g) Whether the discussions of these and other issues in Chapter 3 accurately reflect and clearly communicate the currently available health effects evidence, including important uncertainties, as characterized in the ISA?

The evidence for effects of sub-daily peak exposures to $PM_{2.5}$ are in my opinion undervalued in the PA and ISA. The PA concludes that there is insufficient evidence for consideration of averaging times less than 24-hours. However, this lack of evidence is largely driven by the current form of the $PM_{2.5}$ NAAQS based on annual and 24-hour averages. The epidemiologic studies are largely based on exposure measurement methods which follow the EPA FRMs and NAAQS. Thus few studies have considered sub-daily exposures. Consideration of peak versus 24-hour mean is not equivalent to examining $PM_{2.5}$ associations in the previous hour(s). As the Integrated Science Assessment notes, there are a small number of studies which show increased risk of cardiovascular events (myocardial infarctions and arrhythmias) associated with $PM_{2.5}$ exposures in the previous hours. Likewise, controlled human exposure studies show changes in clinical cardiac indicators after $PM_{2.5}$ exposures of only an hour or less. Wildfire exposures are usually brief (sub-daily) but intense, so understanding the effects of these specific exposures is challenging, but increasingly important.

SCQ 3.4 What are the panel's views on the quantitative risk assessment for $PM_{2.5}$ ***including:***

a) The choice of health outcomes and studies selected for developing concentration-response functions for long and short-term effects?

There are causality determination of "causal "associations for $PM_{2.5}$ with Mortality and Cardiovascular Effects, and "likely causal" for Respiratory, Cancer, and Nervous System effects. However, only a subset of these are included in the risk assessment calculations (see table below).

Health Outcome	Exposure Duration	Causal Determination	Risk Assessment
Mortality	Long & Short	Causal	All Cause Mortality (Long & Short)
Cardiovascular Effects	Long & Short	Causal	Ischemic Heart Disease Death (Long only)
Respiratory Effects	Long & Short	Likely to be Causal	
Cancer	Long	Likely to be Causal	Lung Cancer Deaths (Long only)
Nervous System Effects	Long	Likely to be Causal	

Notably not included:

- Cardiovascular effects (long term) other than IHD mortality, such as cerebrovascular (stroke).
- Any short-term cardiovascular effects (short term), other than IHD mortality
- Respiratory effects either long or short term; mortality or morbidity
- Cancer mortality other than lung cancer
- Nervous system effects (morbidity)

Compare this to the Global Burden of Disease analyses which have developed risk assessment estimates for mortality from All Causes, Ischemic Heart Disease (IHD), Cerebrovascular Events (Stroke), Lower Respiratory Infections (LRI), Chronic Obstructive Pulmonary Disease (COPD) and Lung Cancer

b) The selection criteria for the 47 urban areas and $PM_{2.5}$ ***air quality scenarios analyzed?***

Urban areas were selected for the risk assessment to be in some sense a representative sample of the US population. Three criteria are given for the selection of the 47 urban areas:

- *Available ambient monitors*: *"areas with relatively dense ambient monitoring networks"* This is not defined
- *Geographical Diversity*: *"areas that represent a variety of regions across the U.S. and that include a substantial portion of the U.S. population"* Again not defined and there is not evidence that this was used as selection criteria in that some areas were (e.g. northwest) were explicitly excluded. Note population (>30 yrs) of these areas ranges from ~12 million to ~0.1 million.
- $PM_{2.5}$ *air quality concentrations*: *"areas requiring either a downward adjustment to air quality or a relatively modest upward adjustment (i.e., no more than 2.0 µg/m3 for the annual standard and 5 µg/m3 for the 24-hour standard). In addition, … we excluded several areas that appeared to be strongly influenced by exceptional events."* That is areas above or modestly below the $PM_{2.5}$ NAAQS were included in the initial screen (10/30 criteria). There were multiple adjustments to the air quality data for apparent non-representative values. 56 areas met the initial 10/30 criteria, but 9 (20%) excluded for influence of wildfires (7 areas), high local conditions (Eugene, OR), and "uncertain" projections (Phoenix, AZ).

Overall, these selection criteria are ill defined with post-hoc adjustments that undermine the basis for the selection of theses 47 urban areas. Nevertheless, these urban areas are probably okay as a basis for broad risk assessment. However, they explicitly have excluded consideration of impact of wildfires (which are outside of EPA control), but also local and seasonal sources (wood burning).

c) The hybrid modeling approach used for quantifying exposure surrogates across an area and adjusting air quality for alternative standard levels, as supplemented by interpolation/extrapolation?

Understand that the objective was to provide scaling factor to bring the values at the highest monitor in selected urban areas into compliance with current or proposed alternative standards. However following the logic and process was almost impossible, either in text or appendix. A more detailed flow chart may have been helpful. Not being able to follow undermined my confidence in the calculations.

d) The characterization of variability and uncertainty in the risk assessment?

The characterization of the uncertainties and variability of the risk assessment is ad hoc. Alternative exposure response functions considered, including their individual confidence intervals. However, generally the highest value cited, with no assessment of a central value or range of values across alternative exposure-response functions. Alternative approaches for achieving standards (PM primary and PM secondary) considered, but effectively no consideration of uncertainties in exposure estimates. Recall also that only a subset of health outcomes found to be "causal" or "likely to be causal" are considered, so estimated numbers will be a subset of expected health numbers.

e) The robustness and validity of the risk estimates?

While there are many concerns with the specific approach, the calculated number appear to be of the right order of magnitude. Not clear that the actual numbers contributed to the policy assessment.

SCQ-3.5 What are the panel's views on the draft PA preliminary conclusion that, taken together, the available scientific evidence, air quality analyses, and the risk assessment can reasonably be viewed as calling into question the adequacy of the public health protection afforded by the current primary $PM_{2.5}$ standards?

The scientific evidence from the epidemiologic studies with supporting experimental evidence from controlled human exposure studies and animal toxicology is unambiguous in showing effects of $PM_{2.5}$ at levels below the current primary standards (NAAQS).

The air quality analysis of mean values and distributions of $PM_{2.5}$ values in the key epidemiologic studies, comparing to design values, and examination of "pseudo-design values" addresses some secondary questions in extrapolating from the epidemiologic studies to practical control issues. However, there are some logical flaws in these analyses and the insights are not helpful in drawing conclusions regarding the adequacy of the primary $PM_{2.5}$ standards.

The evidence from the risk assessment, which is built on the epidemiologic evidence and the air quality analyses is confusing and not very helpful.

EPA-6. Chapters 3 to 5: What are the CASAC views regarding the areas for additional research identified in Chapters 3, 4 and 5? Are there additional areas that should be highlighted?

Sub-Daily $PM_{2.5}$ Exposures: The PA concludes that there is insufficient evidence for consideration of averaging times less than 24-hours. However, this lack of evidence is largely driven by the current form of the $PM_{2.5}$ NAAQS based on annual and 24-hour averages. The epidemiologic studies are largely based on exposure measurement methods which follow the EPA FRMs and NAAQS. Thus few studies have considered sub-daily exposures. Consideration of peak versus 24-hur mean is not equivalent to examining $PM_{2.5}$ associations in the previous hour(s). As the Integrated Science Assessment notes, there are a small number of studies which show increased risk of cardiovascular events (myocardial infarctions and arrhythmias) associated with $PM_{2.5}$ exposures in the previous hours. Likewise, controlled human exposure studies show changes in clinical cardiac indicators after $PM_{2.5}$ exposures of only an hour or less. As more continuous PM data becomes available, these sub-daily associations should be examined. Note that wildfire exposures are usually brief (sub-daily) but intense, so understanding the effects of these specific exposures is challenging, but increasingly important.

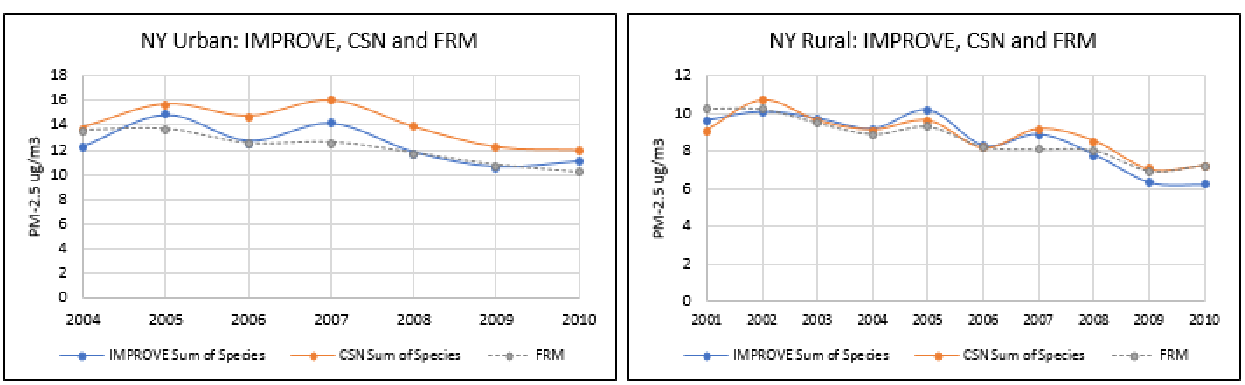
Mr. Henry (Dirk) Felton
October 7, 2019

SCQ 2.1

The draft PA does not provide a clear and concise summary of air quality. When data from different monitoring programs are discussed, inconsistent date ranges are used. The PM data are presented as design values from 2015-2017, ultrafine data are presented for 2014-2015, IMPROVE data are presented from 2004 and 2016 and the analysis on background PM used 2016 IMPROVE data. These data sets from different time periods were then compared to model results for 2011 and source categories from the 2014 NEI. Taking data from different date ranges reduces the validity of the conclusions that can be drawn. For instance, 2016 was a year that included the Fort McMurray wildfire in Alberta Canada. That year should not have been singled out as a representative year to look at background PM.

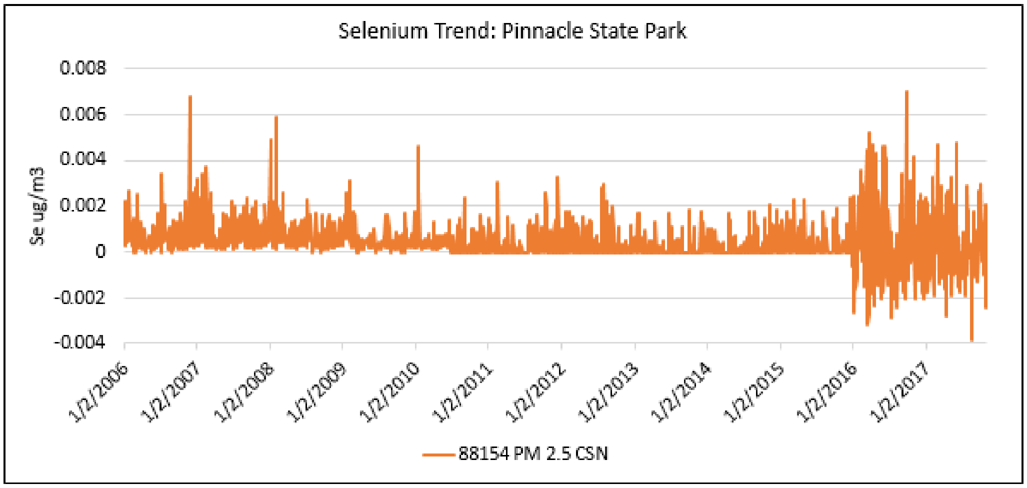
The draft PA provided very little information about the components of $PM_{2.5}$. The National plots only included four species and no elemental data and no mass balance analysis was provided.

The draft PA's summary of air quality should address the shortcomings of the CSN program. This program was originally designed with six objectives linked to assessing $PM_{2.5}$ components over time so States could develop and track SIPs and related control programs. One objective included comparing the mostly urban CSN data with the mostly rural IMPROVE program. Over time, the CSN sampling protocol and the analysis methods for some of the species have been changed to more closely align with the equipment and methods used in the IMPROVE program. These changes have been to the detriment of the State Agencies who need this data to align as closely as possible with the equipment and protocols related to the $PM_{2.5}$ FRM. NYSDEC operated collocated CSN and IMPROVE sites to assess the differences between the programs. The two methods were in better agreement at the rural site where volatile species including OC were lower. At the urban site where the accuracy of the species data were critically important for source attribution, SIP development and control strategy tracking, the CSN results were too high in comparison to the $PM_{2.5}$ FRM.



Data from the South Bronx and from Pinnacle State Park

The CSN program has also been impacted by contractual changes. In late 2015, the CSN laboratory contract was awarded to a different laboratory and this has negatively impacted many of the elemental results. Some low concentration elements have been useful because they can be linked to specific source categories. This data has been used to identify local and out of State source impacts so they can be addressed appropriately. In the plot below, Selenium which has been used to identify coal combustion does not have a useful trend after the change in laboratories.



Selenium CSN Data from Pinnacle State Park

The CSN program is a valuable resource but it has been compromised by competing interests and as a result correction factors have to be applied to various species and some of the elemental data can no longer be used to detect trends. This program needs to be redesigned to make it more representative of the $PM_{2.5}$ in urban areas where ambient concentrations are likely to be closer to the primary NAAQS.

SCQ 3.3

d) Setting a health-based standard that only attempts to limit detrimental health effects for the population within the 25th and the 75th percentile of annual PM concentrations does not represent an adequate margin of safety for at least one quarter of the population. In fact, the admission that the level based on this analysis would only protect a portion of the population against "an array of serious health effects, including premature mortality and increased hospitalizations for cardiovascular and respiratory effects" shows that little attention is paid to susceptible populations and no protections are afforded for health effects short of hospitalization and mortality.

SCQ 3.5

The draft PA addresses each element of the NAAQS individually: indicator, averaging time, form and level. The problem is that the analyses of PM concentration and health effects that accompany each element do not examine each element in isolation. The analyses that accompany the discussion about levels only examine studies that conform to either the averaging time and form of the annual or 24-hour standard. No effort was made to examine health effects resulting from data collected using other averaging times or forms. With this kind of limit: "blinders" on analyses, there is no opportunity to demonstrate the need for a sub daily or alternate form of the standards.

SCQ 3.6

a) The current PM-10 and $PM_{2.5}$ standards are set to protect against respiratory and circulatory system health impacts. Ultrafine particles (UFP) have an additional central nervous system (CNS) health exposure pathway that is not controlled by a standard. A new standard should be set to reduce exposures to higher UFP levels. Some of the largest sources of UFP are combustion sources including stationary and motor vehicles. Motor vehicle emissions can be high from HDD vehicles that have damaged or poorly maintained emission control systems. Vehicle brake and tire wear are also sources that impact most of the population. Setting a UFP standard with a short averaging time would help drive improved controls on sources including HDD vehicles and would reduce exposures in near road communities.

b) The averaging times of the existing PM standards do not adequately protect populations exposed to elevated PM concentrations (UFP, $PM_{2.5}$, PM-10) typically found near roadways during weekday morning commuting hours. These impacts are often the highest exposures in many communities and are more evident near roadways with a higher proportion of HDD vehicles.

The beginning and end times for the averaging time of the 24-hr standard are also not adequate to protect against residential heating and recreational wood smoke impacts. The occurrences of these emissions typically begin in the evening and end in the early morning. The midnight to midnight form of the 24-hour standard effective cuts these impacts into two which in many cases ends up reducing the regulatory impact by averaging additional cleaner hours of two days. Monitoring data have shown that exceedances of the 24-hour standard would be more frequent if the standard were based on noon to noon or on a rolling 24-hour average basis.

f) A lower annual standard does not do enough to reduce the impact from short-term or sporadic sources such as wood smoke from building heating, agricultural burning or industrial activity. Impacts from these sources can have very significant impacts on smaller scales in urban or rural communities. These emissions must be controlled if they impact fewer people just as much as the sources that impact larger scales.

Another disadvantage of lowering just the annual standard is that it may increase the number of times when there is a 10 ug/m3 change in concentration. Health effects have been found to occur when there are 10 ug/m3 changes in air quality in relatively clean and in relatively polluted cities. To prevent these harmful swings in air quality, the daily standard must be lowered in conjunction with or prior to lowering the annual standard.

GC-4

Peaks in background PM are often the result of wildfire emissions or dust storms. These sporadic emissions should not be included in a discussion of peak background PM relative to a NAAQS because these emissions can be excluded from attainment consideration using the exceptional events policy.

Peaks in concentrations resulting from anthropogenic emissions do need to be included in NAAQS data assessments. In urban areas where $PM_{2.5}$ concentrations are closer to current NAAQS, contributions from background PM sources are smaller and less relevant.

Dr. Patrick Kinney
October 8, 2019

Section 3.2 Evidence-Based Considerations

Overall this section is well done. However, I do have a serious concern about the footnote to Table 3-1 on page 3-18. The table lists causality determinations in the 2009 PM ISA and 2018 draft ISA. These provide a central foundation for the entire chapter on primary NAAQS recommendations. The footnote says that the table does not reflect CASAC advice on the draft ISA and that "some or all of these causality determinations could differ in the final ISA." If interpreted literally, this clause opens the door for a complete revision to the evidence on causality which then feeds into the discussions and recommendations regarding the primary NAAQS. This seems like a sort of poison pill for the entire section, which as I said is very well done.

Page 3-61, line 9, and elsewhere in this section. The statements about PM concentrations "around", i.e., "somewhat below to somewhat above" the overall mean observed in the key long-term epidemiology studies is rather vague. I am pleased to see that this notion is made more explicit on the following page, line 7, where there is a suggestion to use the 10^{th} or 25^{th} percentile of the health or concentration distribution to define the lower bound of the data region in which epi results are most precise. These are then plotted in figures 3-7 and 3-8, which is very helpful.

The pseudo-design value analysis starting on page 3-70 provides a useful complement to the previous sections.

Section 3.3 Risk-based considerations

Again, this section is well done, incorporating an appropriate set of inputs and assumptions to examine health outcomes which might occur under a range of assumptions regarding the primary NAAQS.

SCQ-3.2 What are the panel's views on the relative weight that the draft Policy Assessment gives to the evidence-based (i.e. draft PA, section 3.2) and risk-based (i.e. draft PA, section 3.3) approaches in reaching conclusions and recommendations regarding current and alternative $PM_{2.5}$ standards?

Both sets of evidence are given appropriate weight in the draft PA.

Section 3.4 Preliminary Conclusions on the Primary $PM_{2.5}$ Standards

This section accurately recaps and summarizes the evidence, analyses and arguments that were presented in sections 3.2 and 3.3. The draft PA reaches the following appropriate conclusions starting at the bottom of page 3-97.

- There is a long-standing body of strong health evidence demonstrating relationships between long- or short-term $PM_{2.5}$ exposures and a variety of outcomes, including mortality and serious morbidity effects. Studies published since the last review have reduced key uncertainties and broadened our understanding of the health effects 1 that can result from exposures to $PM_{2.5}$.

- Recent U.S. and Canadian epidemiologic studies provide support for generally positive and statistically significant health effect associations across a broad range of ambient $PM_{2.5}$ concentrations, including for air quality distributions with overall mean concentrations lower than in the last review and for distributions likely to be allowed by the current primary $PM_{2.5}$ standards.

- Analyses of $PM_{2.5}$ pseudo-design values additionally support the occurrence of positive and statistically significant health effect associations based largely on air quality likely to have met the current annual and 24-hour primary standards.

- The risk assessment estimates that the current primary $PM_{2.5}$ standards could allow a substantial number of $PM_{2.5}$-associated deaths in the U.S. The large majority of these estimated deaths are associated with the annual average $PM_{2.5}$ concentrations near (and above in some cases) the average concentrations in key epidemiologic studies reporting positive and statistically significant health effect associations.

When taken together, we reach the preliminary conclusion that the available scientific evidence, air quality analyses, and the risk assessment, as summarized above, can reasonably be viewed as calling into question the adequacy of the public health protection afforded by the combination of the current annual and 24-hour primary $PM_{2.5}$ standards.

This material is then followed by a section that presents an alternative, more skeptical, interpretation of the evidence, highlighting uncertainties in biological pathways, potential for public health improvements below the current NAAQS (because accountability studies haven't examined those levels yet), and in risk assessment as a tool. This is a rather extreme interpretation that runs counter to most current scientific views of the available evidence. However, it does provide the administrator considerable scope in evaluating the primary $PM_{2.5}$ NAAQS.

Sections 4.1-4.3 regarding the PM_{10} standard.
Sections 5.1-5.3 regarding the secondary standard.

I reviewed both sections and found both to be well done and to have reached reasonable conclusions. I note that I am not an expert on this literature, so was not in a position to independently evaluate the underlying evidence.

Mr. Richard Poirot
October 7, 2019

Chapter 2

Overall, chapter 2 accurately reflects and clearly communicates air quality information relevant to conducting evidence-based assessments of health and welfare effects, and for conducting risk assessments for evaluating effects of current and alternative NAAQS. Relationships between recent daily and annual PM exposures are clearly presented (for example Figure 2-11). The different hybrid modeling approaches used to estimate exposures are logically derived and clearly described, and information on so-called "background" PM levels are more or less clearly explained. Some interesting results from recent near-road monitoring efforts were presented, although it wasn't clear how/if these results were folded into the hybrid modeling analyses.

Regarding background PM, it's not clear how this information is or will be useful in reviewing and potentially revising the NAAQS. Clearly it could be useful in the implementation phase of the NAAQS (identifying/ getting exemptions for "exceptional" and/or "natural" events). As illustrated, influence of some of these background contributions may be relatively easy to identify and quantify (especially episodes), but I assume we're not saying these background influences don't contribute to health effects (right?). In addition, I think there are likely complex interactions between so-called background influences and "jurisdictionally-controllable anthropogenic sources (see p. 2-49 comment below).

An additional comment on Chapter 2 is that it would be helpful to see some graphic depictions (a few maps and perhaps a time series like Figure 2.6 but for recent years) showing the locations and numbers of the various different $PM_{2.5}$ monitoring techniques/networks (filter FRMs, filter CSN, IMPROVE, continuous (FEM & non-FEM), near-road, etc. A few national maps on this would be useful, as well as a few zoomed-in urban area examples - from some of the cities used in the Risk Assessment (maybe underlain by the hybrid modeling grid). I think this kind of information on monitoring network configurations is especially important given EPA's recent staled reluctance (for secondary PM light extinction NAAQS;

p. 2-2, lines 20-22: I don't think its correct that only a small fraction of coarse mode mass occurs in particles > 10 microns. Much of the coarse mode mass is often > 10 um. See for example Brook et al. (1997), who noted that averaged across 19 long-term Canadian NAAPS sites that "$PM_{2.5}$ accounted for 49% of the PM_{10}, and PM_{10} accounted for 44% of the TSP". This would leave > 4 times more coarse mass in particles >10 um than in $PM_{10-2.5}$.

p. 2-7, lines 7-8: You could also mention ammonium as an important component of $PM_{2.5}$.

p. 2-9, lines 4-8 (similar to above comment): Why not add ammonia to your list of important precursor gasses, instead of just indicating that it "also contributes".

p. 2-49, section 2.4: Here and/or elsewhere when you discuss "natural" vs. "anthropogenic" aerosols, you could add some discussion of PM that results from combinations of natural & manmade sources. For example, emissions from a "natural" dust storm may be enhanced by human actions such as cattle grazing, desert recreation activities, or climate change. "Wildfires" may be started by a careless match, enhanced by historical forest & fire management practices, climate change. "Natural" VOCs may be converted to SOA by reactions with manmade oxidants or through reactions catalyzed by acidic (sulfate) aerosols. Natural sea salt or dust reacts with manmade nitric or sulfuric acids, etc. Sea salt emissions are projected to increase due to climate-driven increases in surface wind speeds. Historical and continuing US emissions represent the largest contribution of any country to the cumulative buildup of global climate-forcing greenhouse gases. Thus, a fraction of transcontinental dust and smoke (considered both "natural" and "non-US") PM reaching the US may have been enhanced by effects our own anthropogenic GHG emissions

Chapter 3
SCQ 3-4
This is not my area of expertise and I defer to other panelists for their thoughts on the quantitative risk assessment. Overall, I found the choices of health outcomes and studies selected fro developing long-term and short-term CR functions reasonable and clearly justified. The selection criteria for included urban areas appear to be similarly logical and clearly described. Variability and uncertainty are clearly characterized, and the results appear to be valid and robust.

I strongly support the hybrid modeling approach as a way of estimating effects that would occur over a range of current and alternative standards (a more realistic improvement over the statistical "quadratic rollback" approach employed several NAAQS review cycles ago). I did find myself curious to better understand the spatial scales, species specifics, and proportionate emissions reductions which ended up being used to meet the various PM concentration thresholds in the different urban areas. Would it be possible to provide some of this information in Appendix C? If not a tabular summary, maybe a few examples could be presented from cities in different regions, showing the spatial scales and absolute reductions (or increases) required of specific primary and secondary emissions species associated with the different PM thresholds evaluated.

Chapter 5
Welfare effects considered in Chapter 5 include those on climate materials and visibility. Some new information is available on climate effects, and while these remain complex, mixed, and uncertain for various PM species, I think a reasonable argument could probably be developed in support of climate-related reductions in black (& brown) carbon concentrations. Some interesting new work quantifying PM materials (soiling) effects on efficiency of solar panels is presented, but does not seem to lend itself to setting a quantitative secondary NAAQS. Relatively little new information is available on visibility effects (although I think some useful recent information on visibility preference indices has been overlooked in the ISA and PAD (more on this below).

SCQ-5.1
The policy questions raised in Chapter 5 relate primarily to visibility. These questions essentially begin with the assumptions that the indicator, level, averaging time and form of the visibility-related PM NAAQS considered (and rejected) in 2012 - are all appropriate, state of the science, and need no further justification or reconsideration. The PAD furthermore jumps immediately to the weakest end (30 DV) of the previously considered 20 to 30 dv range, combined with the weakest (90th percentile) end of the previously recommended 90th to 98th percentile range when considering possible future benefits (of which - Surprise! - there are none). I think all 4 elements of the secondary PM NAAQS considered in the 2012 review need to be reconsidered, justified (if possible) compared to alternatives, and, if warranted, revised.

SCQ-5.2 What are the panel's views of the draft PA evaluation of the currently available scientific evidence with respect to the welfare effects of PM. Does the assessment appropriately account for any new information related to factors that influence:

Quantification of visibility impairment associated with $PM_{2.5}$ and examination of methods for characterizing visibility and its value to the public?
Regarding charge question GC-1, I have concerns with all 4 elements ((indicator, averaging time, level and form) of the secondary PM NAAQS presented for consideration in the Draft PAD document (and rubber-stamped from the 2012 review), and these relate in several cases to information not considered in the ISA.

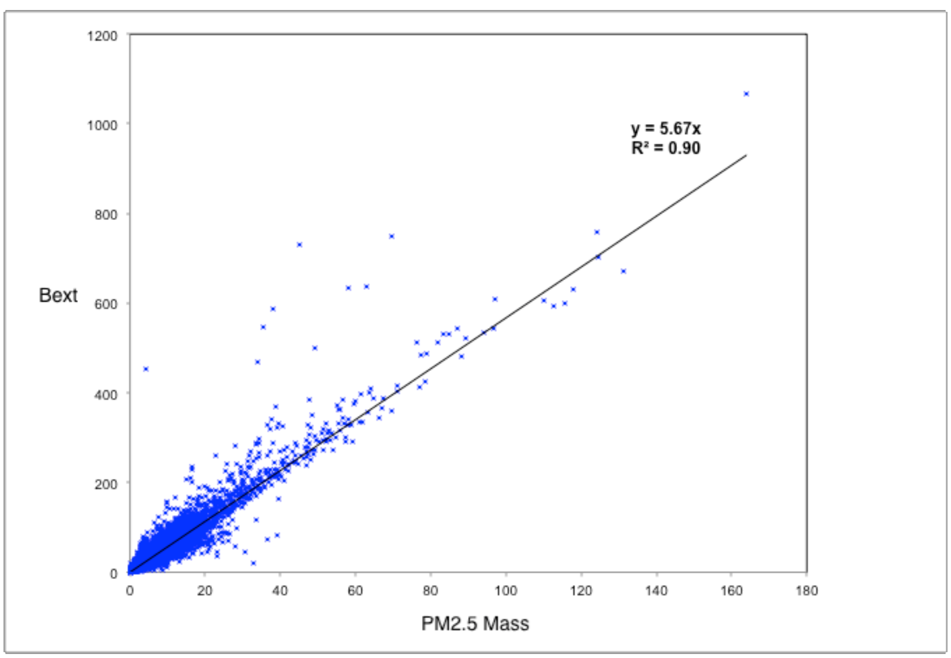
Indicator (reconstructed PM light extinction from 2012 review)
The first PM NAAQS established in 1970 included a separate secondary standard with a PM mass-based indicator (TSP). In subsequent NAAQS reviews completed in 1987, 1997 and 2006, EPA considered, with CASAC support, setting separate, visibility-related secondary PM NAAQS, in each case with a $PM_{2.5}$ indicator (although separate secondary standards were not set after those reviews). In the last review completed in 2012, EPA staff, with CASAC support, considered a different indicator: PM light extinction. During much of that recent review, it was assumed that PM light extinction (or $PM_{2.5}$ light extinction) could and would be directly measured by available continuous methods, such as nephelometer and Aethalometer.

Late in the review, it became clear that the Agency had no intention (resources, will, etc.) of establishing a new national monitoring network, and a fallback methodology was employed to calculate PM light extinction from 1-in-3-day 24-hour filters collected in the EPA STN network and at similar state-sponsored speciation sites using the revised (II) IMPROVE algorithm. This approach takes into consideration the differential densities, size distributions, light scattering and absorption properties and water retention characteristics of different aerosol species. This is basically the method employed to define visibility impairment and track long-term progress toward improving it in Class I National Parks and Wilderness Areas under the Regional Haze Rule. It is not, however, necessarily any better (or as good as) a much simpler $PM_{2.5}$ mass indicator, especially compared to the benefits of using the data from the existing continuous $PM_{2.5}$ monitors in urban/suburban sites.

- The continuous $PM_{2.5}$ network includes 6 times as many sites as the CSN network, providing much better spatial coverage.
- The CSN network samples only every 3rd day, at best, leaving 2/3 of days unmonitored, compared to hourly sampling, every day in the continuous $PM_{2.5}$ network, providing 72x more temporal information (at 6x more sites).
- Filter-based CSN monitoring allows only "24-hour average extinction" estimates. This is not the averaging time over which people perceive impairment. Shorter hourly or 4 to 8-hour (daylight) averaging times would be much more appropriate, especially in urban/suburban areas where light pollution and other factors render night-time PM visibility impairment much less important. Focusing on daylight or mid-day hours would also minimize the importance of RH & speciation, leading to even tighter relationships between actual short-term visibility effects and $PM_{2.5}$ mass data (which are pretty good already). See CASAC recommendations on this from the 2006 review (Hopke et al., 2005).
- While the IMPROVE algorithm - perhaps as enhanced by changes such as suggested by Lowenthal and Kumar (2016) - is "state of the science", it still requires assumptions which are not always well met (the degree of sulfate ammoniation, chemical form(s) of nitrate, the varying relationships between measured OC and POM, etc.) See for example Hand et al., (2019); Preni et al. (2019). Use of 24-hr data also inflates the influence of higher nighttime RH (when urban visibility is least important).

- The filter-based algorithm itself has problems (which appear to be getting worse over time) in reproducing light extinction measured by nephelometry, and conversely, nephelometers have been successfully deployed as $PM_{2.5}$ monitors..
- A good argument can also be made that influence of (naturally) varying RH should be removed from the regulatory metric. Water influence would be minimized by focusing on the (more important) daytime hours. You could also use a fixed, long-term average RH to remove the natural variability from the regulation, or you could impose an RH screen (say eliminating hours with RH < 70%) on the PM data (as is done with urban visibility standards in Phoenix and Denver). Water effects are also decreasing over time as sulfate, nitrate and secondary semi-volatile organics decrease. I don't think you really want the most extreme events driven by extreme uncontrollable variations in RH.
- Use of hourly data would allow eliminating hours with natural impairment (rain, snow, fog).
- Continuous PM data would allow extinction estimates to be publicly reported in near-real-time, rather than waiting for months for the filter data results. Also, since the filter measurements are dependent on multiple samplers at each site, there's more chance of malfunction (on the 1/3 of days when samplers run).
- Use of the continuous PM data for secondary NAAQS regulatory purposes would lead to (needed) closer scrutiny, improved QA and better data quality.
- Light extinction from coarse particles is relatively unimportant in most regions and seasons, and when/where it is important (Southwest, spring), it's often primarily due to natural sources. Alternatively, you could require added use of colocated (or nearest) continuous PM_{10} samplers in areas like the Southwest where coarse particle scattering is important, or set a fine particle NAAQS this time and add a coarse PM component next time.

The figure below is based on all the (unscreened) IMPROVE data from all sites for the 3-year period 2015-2017, limited to sample days when both $PM_{2.5}$ mass and filter-based light extinction estimates are available (about 50,000 sample days).



Similar high correlations have also been observed between continuous $PM_{2.5}$ mass-based $PM_{2.5}$ monitors and nephelometers (or when the continuous nephelometer results are aggregated to 24-hr means - for comparison to filter $PM_{2.5.}$ See for example Chung et al. (2001), Chow et al (2006), Puget Sound (2001), Snider et al., 2015, etc.

The bottom line is that fine mass is a very good indicator of visibility effects, and the small amount of information gained by using speciation filter-based estimates is way more than offset by the spatial, temporal information and visibility relevance that would be gained using continuous $PM_{2.5}$ monitors and a sub-daily daytime averaging time. Please note the CASAC comments on secondary NAAQS from the review completed in 2006, for example: Hopke et al. (2004), page 9 and pages B9-B26.

If you really want to keep the light extinction indicator, use the filter-based speciation data to calculate regional monthly or seasonal species composition + f(RH) factors to adjust the continuous $PM_{2.5}$ data to (slightly) better extinction estimates - which could then be considered on a sub-daily basis, much more relevant to human perception, and could be publicly reported from a much larger network in near-real time. I think a simple sub-daily $PM_{2.5}$ mass indicator which intentionally limits the influence of naturally varying RH on the regulatory metric is a much better choice for an indicator. If the Agency wants to persist in advocating continued use of the every 3rd day 24-hour, filter-based reconstructed light extinction indicator, it needs to justify why it thinks it has a superior indicator. I don't think it can.

Averaging Time (24 Hours from 2012 PM NAAQS)
As indicated above, once the decision was made that the PM light extinction indicator introduced in the 2012 review would not be measured directly and continuously, a fallback method was proposed to calculate PM light extinction based on every 3rd day 24-hour filter sampling. This limits the averaging time to to no shorter than 24 hours, which is not the time frame over which visibility impairment is perceived. It's also especially inappropriate in urban areas where visibility during daylight hours is much more important (and is also characterized by lower RH levels - reducing the small difference between PM mass and light extinction).

EPA's Final Staff Paper (US EPA, 2005) from the 2006 PM NAAQS review (which recommended a sub-daily $PM_{2.5}$ mass indicator in the range of 20 to 30 ug/m^3 for a secondary PM NAAQS) stated:

> In considering appropriate averaging times for a standard to address visibility impairment, staff has considered averaging times that range from 24 to 4 hours, as discussed in section 6.2.3. Within this range, as noted above, correlations between $PM_{2.5}$ concentrations and RE [reconstructed extinction] are generally less influenced by relative humidity and more consistent across regions as the averaging time gets shorter. Based on the regional and national average statistics considered in this analysis, staff observes that in the 4-hour time period between 12:00 and 4:00 p.m., the slope of the correlation between $PM_{2.5}$ concentrations and hourly RE is lowest and most consistent across regions than for any other 4-hour or longer time period within a day (Chapter 6, Figure 6-4). Staff also recognizes that these advantages remain in looking at a somewhat wider time period, from approximately 10:00 am to 6:00 pm. Staff concludes that an averaging time from 4 to 8 hours, generally within the time period from 10:00 am to 6:00 pm, should be considered for a standard to address visibility impairment.

It can also be noted that the quality of the continuous $PM_{2.5}$ mass data has improved considerably over the past 20+ years, providing greater confidence in its accuracy over shorter averaging times. The Agency should consider the many benefits of using the continuous $PM_{2.5}$

data as the measurement basis for a sub daily 4 to 8 hour daylight averaging time. This could be combined with a continuous $PM_{2.5}$ mass indicator, or regional & seasonal generic species composition and f(RH) factors could be developed to convert the mass to estimated extinction. Either way, the data could be reported in near-real time, and would relate more directly to the human perception of impaired visibility.

Level (20 to 30 dv from 2012 review)

While the previous 2011 PAD recommended a range of 20 to 30 dv as an appropriate level, the Administrator (sort of) picked the upper end, before concluding that such a NAAQS wouldn't do much good. The current PAD simply starts with this upper end (30 dv) as if this were a logical, technically-supported absolute definition of "acceptable visibility or adverse effects. It's not.

In a previous review of the draft ISA, I noted that the ISA had neglected an important recent meta-analysis of visibility preference studies by Bill Malm (Malm et al. 2011, 2018 and Malm, 2016) which could support an alternative way of defining adversity based on geographical differences in distant landscape features. My earlier comments on this omission in the ISA are pasted below:

> A second general criticism of this brief summary - as well as with the more detailed Chapter 13 discussion of visibility - is the absence of discussion of recent work on visibility preference indicators developed by William Malm over the past several years (Malm et al. 2011, 2018 and Malm, 2016). His meta analysis of multiple available visibility preference studies (in many different kinds of locations) noted that "unacceptable" levels of visibility impairment occurred at different extinction levels in different areas, but that in any area, when the more-distant visible landscape features nearly disappear - which occurs at apparent contrast levels of about 0.02–0.05 - the haze level became unacceptable to about half of the participants in each study area. This has important implications for the potential setting of PM visibility standards at nationally consistent contrast levels which are geographically variable with changing distant landscape features. It would b a relatively straightforward GIS exercise to characterize distances to prominent landscape features in population centers throughout the country and then use $PM_{2.5}$ -based extinction estimates to calculate contrast levels for those landscape objects to determine the extent to which visual air quality is (or is not) considered acceptable in each of those areas.
>
> There appears to be a reference to Malm's work in the executive summary: "There have been no recent visibility preference studies; however, a recent meta-analysis demonstrates that scene-dependent haze metrics better account for preference compared to only using the deciview scale as a metric." However, any discussion of this recent work seems to be missing from the Integrated Synthesis or Chapter 13. Section 13.2.5 on "human perception of haze and landscape features" heavily emphasizes the divergent results in different visibility preference studies in areas with (or using photographs showing) different landscape features, when visual air quality is expressed as light extinction (deciviews). It concludes with:
>
> "There is little new published information regarding preference levels in the U.S. The single new study by Smith (2013) was an investigation of "framing bias" in preference studies that can potentially occur because preference levels are chosen in part based on experimental variables such as number of photographs shown or range of the range of dv levels participants are shown when asked to state a preference about whether scenes in photographs are acceptable."

> This disregards important new work in this area, which clearly shows a convergence of results across many different urban areas when the visual air quality is expressed in terms of the contrast of the most distant landscape features. Another important recent related technological development is the ability to incorporate clouds into the Winhaze model - developed by John Molenar (Molenar and Malm, 2012). For cities in relatively flat terrain which lack distant landscape features, clouds often are the most distant scenic attribute. As they begin to disappear, viewers tend to find the degradation of visibility unacceptable, at lower levels of light extinction than they would viewing cloud-free scenes. Some discussion of this work, implications and potential future applications is warranted in chapter 13.

Please see Figures 4 and 5 from **Malm et al., 2019,** which illustrate the consistency across may varied study areas using the apparent contrast of distant landscape features (Figure 5) vs. using absolute extinction (Figure 4 - similar to Figure 5.2 in the PAD - except Malm includes more studies). Note the convergence in Figure 5 of 50% acceptability across many diverse sites at an apparent contrast threshold of about -0.04. Stated in simpler terms, people tend to find decreased visibility unacceptable when prominent, distant landscape features begin to disappear.

If this kind of approach were applied across multiple urban/suburban areas throughout the country, it would be clear that people in many diverse regions would likely find visibility impairment of 30 dv to be unacceptable. The Agency should consider using this apparent contrast threshold as a basis for setting a consistent national standard which could vary geographically depending on local scene characteristics. This would be a similar concept to what the Agency considered in the last review of secondary SOx + NOx NAAQS, in which the varying biogeochemical features of local eco-regions were incorporated into the proposed standard.

Form (90th Percentile from the 2012 review)
The 90th percentile is not supported in the PAD or ISA. Its just repeated from the last review cycle (where it was never justified either). It was simply a way for a secondary NAAQS - considered at the most lenient end of the staff-recommended 20 to 30 Dv range - would have little to no benefit over the primary standard. The forms of the various secondary standards that have been considered/ recommended by CASAC over the years has varied widely: not to be exceeded more than 1 day/year (1971), 3-month seasonal mean averaged spatially over multiple years (1987), 98th percentile averaged over 3 years (1997), 92nd to 98th percentile, 3-year average (2006), and 90th to 98th percentile (2012).

With the exception of 1971, when a separate secondary PM standard was set, the secondary NAAQS considered in all subsequent reviews were rejected for various “reasons” (see: Poirot, 2011). In the 2012 review, the Administrator selected the (most lenient) 90th percentile combined with the weakest level (30dv) before concluding that this combination really wouldn’t have much incremental benefit over the primary. The only stated justification was that the Regional Haze Rule is focused on the haziest 20% days, and that the 90th percentile - the midpoint of the haze range - would be consistent. The average of haziest 20% days is closer to the 92 percentile (considered as the low bound in 2006 for that stated reason).

More importantly, this is a false equivalency. The focus in the Haze Rule is specifically on <u>improving</u> conditions on these worst days. The use of a similar percentile as a NAAQS form has exactly the opposite effect - of completely <u>ignoring</u> the worst visibility days, exculpating them from any consideration of improvement. Visibility could be worse, or much worse than 30 dv on 36 days each year, but people only find it objectionable when this happens 37 or more days per year (averaged over 3 years). This is not logical, and no other justification is provided in the PAD or ISA.

SCQ-5.3 What are the panel's views of the draft PA preliminary conclusion that the currently available scientific evidence does not call into question the protection afforded by the current secondary PM standards against PM welfare effects and that it is appropriate to consider retaining the current secondary PM standards without revision?

As indicated above, I have criticisms of all elements of the secondary NAAQS which was considered (but ultimately rejected by the Administrator). I also don't think current secondary PM NAAQS provide adequate protection against adverse visibility effects on public welfare. To illustrate the visual air quality effects of the current 24-hour NAAQS, please see Figure 2 (E, F and H) from Poirot (2011) showing a clear day view from Denver which has been modified by a model called WinHaze developed by John Molenar at Air Resource Specialists and now available on-line at: http://vista.cira.colostate.edu/Improve/winhaze/ .

WinHaze models the visual effects of various mixes of pollutant species, concentrations and RH. The upper left (E) and right (F) images show the visual air quality at 40% RH (right) and 70% RH (left) at a PM concentration (composed of ammonium sulfate and ammonium nitrate) equal to the current $PM_{2.5}$ primary and secondary 35 ug/m^3 standard. Note that these also illustrate the approximate high bound (E - 31 dv) and mid-range (F - 25 dv) of the 20 to 30 dv range recommended for consideration in the 2011 EPA PM PAD. The upper bound (30 dv), with a 90th percentile form was considered and rejected by the Administrator in 2012, and is being put forward on the current PAD and again found to be not much of an improvement over the current NAAQS. Is this an adequate level of visibility protection?

Similarly, the current annual secondary PM NAAQS at 15 ug/m3 is weaker than the primary, and therefor protects nothing, nor has any scientific justification been provided

References

Brook, J.R., T. F. Dann & R. T. Burnett (1997) The Relationship Among TSP, PM10, $PM_{2.5}$, and Inorganic Constituents of Atmospheric Participate Matter at Multiple Canadian Locations, *J. Air & Waste Management Assn,* 47:1, 2-19, DOI: 10.1080/10473289.1997.10464407

Chow,J.C, J. G. Watson,.D.H. Lowenthal, L.-W. A. Chen, R. J. Tropp, K.Park and K. A. Magliano (2006) $PM_{2.5}$ and PM10 Mass Measurements in California's San Joaquin Valley, *Aerosol Science and Technology,* 40:10, 796-810.

Chung, A., D. P.Y. Chang , M. J. Kleeman , K. D. Perry , T. A. Cahill , D. Dutcher , E. M. McDougall & K. Stroud (2001) Comparison of Real-Time Instruments Used To Monitor Airborne Particulate Matter, J*ournal of the Air & Waste Management Association*, 51:1, 109-120, DOI: 10.1080/10473289.2001.10464254

Hand, J.L., A.J. Prenni , B.A. Schichtel, W.C. Malm, J.C. Chow (2019) Trends in remote $PM_{2.5}$ residual mass across the United States: Implications for aerosol mass reconstruction in the IMPROVE network, *Atmospheric Environment* 203: 141-152.

Hopke, P.K. et al. (2004) CASAC letter to EPA Administrator Leavit, EPA-SAB-CASAC-04-004, https://nepis.epa.gov/Exe/ZyPDF.cgi/P1000WRB.PDF?Dockey=P1000WRB.PDF

Malm, W.C., Molenar, J.V., Pitchford, M.L., Deck, L.B. (2011) Which visibility indicators best represent a population's preference for a level of visual air quality?, Paper 2011-A-596-AWMA, Air & Waste Management Asson. 104th Annual Conference, Orlando, FL, June 21-24, 2011.

Malm, W. C. (2016) Visibility: the seeing of near and distant landscape features, Elsevier, Inc., ISBN: 978-0-12-804450-6

Malm, W., B. Schichtel, J. Molenar, A. Prenni, and M. Peters (2019). Which visibility indicators best represent a population's preference for a level of visual air quality?. *Journal of the Air & Waste Management Association.* 169(2):145-161. Epub 2018 Dec 6.

Molenar, J. V., & Malm, W. C. (2012). Effect of clouds on the perception of regional and urban haze. Paper presented at the Specialty Conference on Aerosol and Atmospheric Optics: Visibility and Air Pollution, Whitefish, MT.

Poirot, R.L. (2011) Will the cycle be unbroken or will we see a secondary PM NAAQS someday? *EM,* AWMA, Sept. 2011, http://pubs.awma.org/gsearch/em/2011/9/poirot.pdf

Puget Sound Clean Air Agency (2001) PM data comparison: the Puget Sound experience, (https://www3.epa.gov/ttnamti1/archive/files/ambient/samwg/puget.pdf)

Prenni, A.J., J.L. Hand, W.C. Malm, S. Copeland, G. Luo, F. Yu, N. Taylor, L.M. Russell, B.A. Schichtel (2019) An examination of the algorithm for estimating light extinction from IMPROVE particle speciation data, *Atmospheric Environment,* Volume 214, 116880.

Snider, G. et al., (2015) SPARTAN: a global network to evaluate and enhance satellite-based estimates of ground-level particulate matter for global health applications, *Atmos. Meas. Tech.*, 8, 505–521.

US EPA (2005) Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Scientific and Technical Information, OAQPS Staff Paper, EPA-452/R-05-005, https://www3.epa.gov/ttn/naaqs/standards/pm/data/pmstaffpaper_20050630.pdf

US EPA (2011) Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standards, EPA 452/R-11-003, https://www3.epa.gov/ttn/naaqs/standards/pm/data/20110419pmpafinal.pdf

Dr. Ronald Wyzga
October 7, 2019

Chapter 1 – Introduction: To what extent does the CASAC find that the information in Chapter 1 is clearly presented and that it provides useful context for the review?

The Chapter is clearly written, but it omits key factors that set the context for this review. First of all, it does not indicate the differences in the overall review process for PM in this review as opposed to previous reviews. Secondly, there was limited review of the ISA with only one draft reviewed despite the comments made on the first draft. Thirdly, there is no formal risk and exposure assessment as has been included in previous reviews. Finally, the content of this chapter is dependent upon the science and conclusions of the ISA. Only a draft version is available; the final version is planned for release in December 2019. Given the uncertainty about the content of this document, it makes it difficult to make this document at best provisional and subject to change given changes in the ISA. This Chapter needs to recognize these factors and indicate how the overall process will accommodate them.

Chapter 3 – Review of the Primary $PM_{2.5}$ **Standards: What are the CASAC views on the approaches described in Chapter 3 to considering the** $PM_{2.5}$ **health effects evidence and the risk assessment in order to inform preliminary conclusions on the primary** $PM_{2.5}$ **standards? What are the CASAC views regarding the rationales supporting the preliminary conclusions on the current and potential** $PM_{2.5}$ **standards?**

SCQ-3.1 Does the panel find that the questions posed in this chapter appropriately reflect the important policy-relevant issues for the $PM_{2.5}$ review? Are there additional policy-relevant questions that should be addressed?

> The content is based upon a draft ISA; it is unclear whether a final ISA would influence the discussions and conclusions of this chapter. By and large the questions addressed are reasonable. I would have like to have seen more discussion of PM components other than ultrafine particles. Although virtually all PM components have been shown to have some adverse health impacts, there are some differences among major components for both respiratory and cardiovascular endpoints. Although these differences would not change the PM indicator, they are noteworthy and could help inform risk managers about the need to consider all major PM components in achieving compliance, I base my conclusions on two relatively recent reviews in which I was involved. A comprehensive review of the literature for both short-term and long-term studies found that different components were associated with respiratory and cardiovascular endpoints; moreover, although no major components of PM were exonerated, there appeared to be greater and more associations with organic particles than with other components. See A. C. Rohr and R. Wyzga . Attributing health effects to individual particulate matter constituents. Atmospheric Environment 62 130-152 2012 and R.E. Wyzga and A. C. Rohr. Long-term Particulate Matter Exposure: Attributing Health Effects to Individual PM Components. . J of the Air and Waste Manage. Assoc. 2014.

SCQ 3.2 What are the panel’s views on the relative weight that the draft Policy Assessment gives to the evidence=based (i.e., draft PA Section 3.2) and risk-based (i.e. draft PA, section 3.3) approaches in reaching conclusions and recommendations regarding current and alternative PM 2.5 standards?

I like the fact that two approaches were considered; the conclusions for each were similar which adds strength to an overall conclusion. Both approaches clearly indicate that the current standard is not protective. These sections do not consider all studies covered in the iSA. Greater justification of the studies considered need be incorporated into the PA.

SCQ 3.3 What are the panel's views on the evidence-based approach, including:

a) The emphasis on health outcomes for which the draft ISA causality determinations are "causal" or "likely causal"?

I have no problem with considering the adverse health ese two categories. It should be noted that consideration of the "causal" and "likely causal" categories will most likely result in standards that are protective of other categories. To the extent that this may not be true, some indication could be useful.

b) The identification of potential at-risk populations?

The draft PA rightly indicates that very large subpopulations are at risk. Greater specificity is not necessary.

c) Reliance on key multicity epidemiology studies conducted in the US and Canada for assessing the PM 2.5 levels associated with health effects?

There should be greater discussion about how the results might change if a broader set of studies considered in the ISA were included here.

d) Characterizing air quality in these key studies using two approaches: the overall mean and 25 th /75 th percentiles of the distribution and the "pseudo design value" reflecting a monitor with the highest levels in an area?

The approach is reasonable although there should be some discussion about the nature of the overall statistical distribution; this may be covered in Chapter2, which I have not yet reviewed.

e) The preference for continuing the use of an annual PM 2.5 standard as the principle means of providing public health protection against the bulk of the distribution of short- and long-term PM 2.5 exposures?

If the analysis were the other way around, would it be as useful? My concern is that some extreme events could possibly alter some of the assumptions between long-term and short-term air quaiity measures.

f) The draft PA conclusions on the extent to which the current scientific information strengthens or alters conclusions reached ion the last review on the health effects of on the health effects of $PM_{2.5_2}$?

I agree with the conclusions.

g) Whether the discussions of these and other issues in Chapter 3 accurately reflect and clearly communicate the currently available health effects evidence, including important uncertainties as characterized in the ISA?

Without seeing the final ISA, it is difficult to evaluate this question. This chapter considers a subset of studies covered in the current iSA; it would be helpful to explain further how the subset was chosen and what would be the impact of considering a wider set of studies.

SCQ 3.4 What are the panel's views on the quantitative risk assessment for $PM_{2.5}$ including:

a. The choice of health outcomes and studies selected for developing concentration-response functions for long and short-term effects?

I would like to see greater explanation of how the selected studies were chosen, and what the likely impact would be if additional studies were chosen as well. I was struck by the fact that the studies that used modeling as opposed to monitoring to estimate PM exposures appeared to give slightly different results. I would like to see some discussion of this. Is it because different geographic regions were considered or some other reason?

b. The selection criteria for the 47 urban areas and $PM_{2.5}$ air quality scenarios analyzed?

No problems here.

c. The hybrid modeling approach used for quantifying exposure surrogates across an area and adjusting air quality for alternative standard levels, as supplemented by interpolation/extrapolation?

It seems reasonable

d. The characterization of variability and uncertainty in the risk assessment?

Again if additional studies were considered, would the results and their variability change much?

e. The robustness and validity of the risk estimates?

I would like to see more discussion of the differences seem in those studies that considered modeling as opposed to monitoring to estimate PM levels.

SCQ-3.5 What are the panel's views on the draft PA preliminary conclusion that, taken together, the available scientific evidence, air quality analyses, and the risk assessment can reasonably be viewed as calling into question the adequacy of the public health protection afforded by the current primary $PM_{2.5}$ standards?

I agree.

SCQ-3.6 What are the panel's views on the conclusions in the draft PA regarding developing potential $PM_{2.5}$ alternative standards with respect to:

a. The preliminary conclusion that the available information continues to support the $PM_{2.5}$ mass-based indicator, remains too limited to support a distinct standard for any specific $PM_{2.5}$ component or group of components, and remains too limited to support a distinct standard for the ultrafine fraction?

The issue here remains tied to the ISA. I agree that no major constituent of PM is exonerated, but the draft ISA, in my opinion, does not fully discuss the relative roles of major constituent categories. See my comments with regard to charge question 3.1.

b. The preliminary conclusion to retain the annual and 24-hour averaging times?

Reasonable ,but is should be pointed out that most studies make use of commonly reported air quality measures. Further research to indicate whether other averaging times would be preferred is lacking.

c. The preliminary conclusion that it is appropriate to consider retaining the forms of the current annual and 24-hour $PM_{2.5}$ standards, in conjunction with revised levels?

d. The preliminary conclusion that the range for alterative levels for the annual $PM_{2.5}$ standard should begin below 12 ug/m^3 and extend as low as 8 u/m^3?

Reasonable

e. The possible rationales for alternative annual $PM_{2.5}$ levels of 12, 10, and 8 ug/m^3?

Reasonable, but I would like to see further discussion of why the Canadian studies and those studies which used modeled air quality data appear to give different results.

f. The preliminary conclusion that, in conjunction with a lower annual standard intended to protect against both short- and long-term exposures, the evidence does not support the need for a revised level for the $PM_{2.5}$ 24.-hour standard?

I worry about this. The arguments supporting this position are not crystal clear to me and all of the assumptions therein need be clearly articulated.

g. The discussion of an alternative approach to lower the level of the 24 hour standard to 30 ug/m^3 to provide increased protection for both short- and long term exposures?

I liked this.

Chapters 3 to 5: What are the CASAC views regarding the areas for additional research identified in Chapters 3, 4 and 5? Are there additional areas that should be highlighted?

The current review must be based upon existing information; however, there are several areas that could inform future reviews of the standard and help reduce some of the uncertainties associated with this process.

I believe that future research should include the following:

- More detailed measurement of PM components; in particular, more detailed measurements of organic components. Several studies have suggested that some organic components may be of greater health concern than others. EC and OC are catchall categories defined by a measurement technique. Availability of such measurements would facilitate their use in future epidemiological studies.

- Research should also continue to define in more detail the physiological bases for adverse health responses to PM and its components. It may be that different components are associated with different components. If so, consideration of components may provide a more precise understanding of the biological basis for observed responses in epidemiological studies.

- Alternative exposure metrics need to be explored. How important are peak exposures as opposed to average exposures in explaining observed health responses? What is the appropriate time average for peak exposures? Do current average measures adequately limit exposures to peak levels? Is the relative change in exposure important; research needs to consider the issue of delta exposure. How important are past exposures in explaining responses to current levels; indeed the correct question to ask is what are the impacts of current exposures given past exposures? This is particularly importance when health outcomes, e.g., cancer, develop over an extended period of time and when cross-sectional designs are considered. These designs compare exposures and health responses across geographic entities. Although there are changes are changes in air quality over time, the relative ordering of air quality across geographic entities changes minimally. What is the latency of response? Tied to this is the issue of cumulative exposure, which should be examined.

- Consideration of the NAAQS for the coarse fraction of PM is limited because measurement of the coarse fraction per se is limited. There are studies, especially considering asthmatic response, that report significant associations with PM_{10} but not $PM_{2.5}$. Statistical and oher phenomena could explain these results, but they could also suggest that coarse PM, independenty of fine PM, may be of health concern. More research on the relationship between asthmatic and other respiratory responses and coarse PM is needed.

- Health research tends to be focused on one pollutant at a time even when several pollutants are measured, but they are most often considered independently. How important is joint exposure to more than one pollutant in influencing health response? Is sequencing of exposures important?

- People spend more of their time in indoor environments. Indoor PM levels can be high in these environments? How important are these? If they are not as important, why? What is the health impact of joint indoor and outdoor exposures? Are health responses to outdoor PM levels greater when indoor levels are high?