Responses to CASAC Questions on the Draft PM PA from Consultant Dr. Lorenz Rhomberg

Questions from Dr. Cox

Dr. Cox's questions do not seem to be designed to be answered individually. Together, they present a set of concerns about causal interpretation of associations and the consequent impact on using association measures as indices for the amount of risk reduction that might accompany specific changes in PM measures that could result from altered standards and subsequent controls. Accordingly, I have answered the set of questions in one answer.

There are several aspects to be addressed. First is the purely technical question of whether Dr. Cox's expressed concerns have merit – are the difficulties in interpretation he raises indeed present? Second, if so, what as a technical matter can be done to alter the analysis so as to avoid or at least minimize the impact of these concerns as they might affect judgments about the impact of changes in PM exposures. Third, in view of the limits to what can be done in this regard, how can achievable analyses of observational data be used to address the fundamental objective of the whole document's analysis: to inform judgments by the Administrator as to what PM standards – and in particular, what changes to those standards – would achieve the objective of protection of public health with an adequate margin of safety. As noted on p.3-15 of the Policy Assessment, "'adequate margin of safety' is meant to address uncertainties associated with inconclusive scientific and technical information."

To start with the first question, Dr. Cox is correct that the measures of association developed in modeling observational studies are not directly measures of causative actions by the factors measured in individual variables, including variables describing PM exposures. They are contingent on the whole array of conditions, influences (known and unknown), as they vary person-by-person and place-by-place and time-by-time in their antecedent states and combinations as potential contributors to the effect or modifiers of the impact of still other factors. (All these are in addition to measurement errors in the variables used.)

The data going into a regression model represent a snapshot of the overwhelmingly complex state of the world as it was upon data collection, with the model used being an attempt to identify the major patterns of association as they were then seen. Of course, the regression may include covariates that aim to "adjust for" the contribution of some factors, but even these adjustments will be contingent on the particular states of the array of the covariate factors at the times of observation. They also depend on the chosen methods for adjustment. They necessarily make hypothetical decisions about the functional form of covariate contribution, how measured averages over space, time, and individuals succeed in representing average impacts, whether or how the adjustment factors covary, and whether they influence one another's' impacts in ways that can be represented by simple model equations. In short, the fitted model represents, in effect, an assertion that the chosen functional forms for interactions among variables are exactly correct, that the way the variables are measured (and how they average over the complexity of observed instances) completely capture the effect (irrespective of unmeasured patterns of underlying variability), and that all underlying variability, unmeasured influences, and contingencies among varying states can be lumped together in a simple unbiased "error" term that needs to represent not only measurement error but also the influence of all the unmeasured contingencies and interactions among particulars.

All this is an unavoidable aspect of observational studies. All one can ask for is that the asserted functional structure of the model is reasonable in view of our broader understanding of what kinds of

causative processes might operate (or be of interest), that the simplifications of the actual causal nexus are not obviously misleading or ignoring known or plausible important influences, and (to speak directly to Dr. Cox's questions) that we do not over-interpret the functional forms as actual measures of causal impacts without considering the pitfalls and limits to such interpretation.

In particular, as Dr. Cox notes, one cannot simply change the value of one variable and suppose that the model's result represents what would be expected in a real setting if that variable had the different value. To do so presupposes that the functional form of the model has captured all influences correctly (and not simply in an average way), and that all the other variables would indeed be unchanged. This is different from changing the value and allowing for the corresponding changes in other variables that might result from the same intervention (or from patterns of interactions among factors). And these are different from asserting that the parameter characterizing the functional dependence of the outcome on a variable is a measure of that variable's causal effect, much less a measure of how a different level of that variable would exert an altered causal effect.

In short, Dr. Cox's concerns have technical merit in terms of demonstrating how translating measures of association can be misleading if interpreted as generally applicable measures of the degree of causal effect imparted by the factor, and how generalizing from the world-state measured in the study to other states entails buying in to the actual truth of the simplifying assertions about functional relationships, lack of interactions, and sufficiency of averaging assumptions that are entailed in any regression approach to observational data.

An important part of the interpretation of causality arguments is to consider the plausibility (in view of our wider biological knowledge and other kinds of data) of the operation of some of the potential complications. For instance, one can note (as I have done) that using group averages for factors that vary among individuals makes assertions that the average effect of the variations is given by the average of the variations themselves. One can sometimes judge which potential complications are serious by using outside information on such things as age dependence, dependence of effects on time patterns of exposure, and so on to judge whether an exposure measure that averages these things out could be potentially misleading.

Another important consideration is confounding factors. As one not steeped in the air pollution epidemiology literature, I was surprised to see that many PM studies do not account for the levels of other air pollutants. Since several important pollutants are markedly correlated in their levels across time and space, and since, when each of these is studied alone, associations with a similar set of outcomes is asserted, it would seem necessary that a study not controlling for copollutants would overestimate the effect attributable to the measured one. If the pollutants act independently, then one is "explaining" the same interlocality variation in endpoint levels as attributable to each one if studied alone, overcounting the total effect. If they interact (additively or synergistically), then the ability to sort out contributions is further complicated and attribution to any one alone is again a significant overestimate.

The second of my questions noted above is what can be done in terms of technical analytical approaches to ameliorate this situation. Dr. Cox names several candidate approaches in his questions (specifically, in his Question 16). I am somewhat familiar with a few of these, but I do not feel able to comment in detail, especially about how well the approaches would resolve the issues for the currently applicable data. I think one would need to try them out to see.

I have an overarching observation, however. As I have set out above, the problem arises because any realized dataset is the product of myriad factors, measured and unmeasured, known or suspected or

unrecognized, that have all contributed in a complex nexus of interactions, causes, prior states, and influences. This array will never be completely specifiable, and in any case, any application of the findings to another time or place will entail a different set of such influences, also not fully specifiable. In short, one cannot generalize an empirical finding as such; one can only hypothesize what abstraction of the particular case might be generalizable and then (1) use whatever means at hand to evaluate how compelling that hypothetical generalization might be, and (2) apply it to make predictions that nonetheless need to be couched in examination of the soundness and sufficiency of the asserted generalizable assertions.

This applies to any deeper analysis of causal patterns. The proffered methods in Dr. Cox's Question 16 might aid in finding bases for stronger hypothesizing of causal influences, but any such effort will ultimately be less than fully sufficient, since any feasible explanatory model must necessarily include unverified assertions about functional relationships among variables (that may be more complicated than is feasible to estimate from real data), about the sufficiency of measurements that average over time or members of populations, *etc*. That is, the challenge of near-infinite complexity simplified for tractability and for focus on main effects is still an issue. Therefore, what can be hoped for is improvement in the ability to hypothesize causal relationships with greater dependability, but with no absolute solution to the challenges.

The third major question I set out at the start of this comment is how this state of affairs should affect how we interpret the association relationships observed as we grapple with finding a basis for projecting what risk changes might arise from alterations in exposures.

It should be stated at this point that, just because it is difficult to rigorously identify and pin down a measured causal effect, this does not mean that there is no such causal effect. It would not be prudent (or consistent with the CAA mandates) to reject a causal effect simply because it cannot be unambiguously measured and characterized. There is no basis to default to a conclusion of no causation unless a rigorous demonstration of its particulars is at hand. We have overall patterns of outcomes in many studies that certainly suggest causative roles for PM in risks of health impacts, and these patterns are repeated with some level of consistency across studies.

It is important to remember that, if one denies the causal influence of a variable (owing to the difficulty in unambiguously demonstrating it or measuring it), then one is faced with proposing some alternative hypnotical causes for the patterns that were found in the data. That is, if PM does not increase mortality as a function of the levels prevailing in each city, but there are mortality differences among the cities, then some other factor must operate to produce the observed differences. The fuller picture of attribution of causality is not just the ability unambiguously to demonstrate and measure it, but also the comparative plausibility of alternative explanations of the patterns that are actually observed. Something must be causing the patterns, and one asks if it is more plausible to accept the causal role of the variable being examined or to assert some one among a set of conceivable alternatives, with evidence for and against the alternatives needing to be weighed.

The analogy is less to a significance test (if the evidence is not strong enough, one defaults to the null) and more like a Bayesian or Likelihood approach (what causal model is most plausible, measured by how likely the results would be under each competing hypothetical attribution of the reasons behind the observed patterns. This is an instance of the larger weight-of-evidence methodology issue, and I refer the CASAC to some publications of mine that set out the above line of argument (e.g., Rhomberg LR. 2014. "Hypothesis-based weight-of-evidence: An approach to assessing causation and its application to regulatory toxicology." Risk Analysis 35(6):1114-1124).

I would propose that it is useful to separate the following questions:

1. What is the basis for asserting that an exposure has some causative effect on an outcome of interest? (This is the qualitative question of whether any causal influence exists.)
2. What is the basis for asserting that such a causal effect has actually operated so as to affect the magnitude of the outcome of interest in a particular setting that has been observed? (This is a retroactive and case-specific assertion of the qualitative operation of the cause.)
3. What is the basis for measuring the degree of effect in such an observed setting? (That is, what change in the outcome measurement is being attributed to the causal factor as manifested in the already observed dataset, *i.e.*, that particular instance of causation?)
4. What is the basis for projecting the degree of change in an outcome of interest as a result of a particular change in the causal variable? (This is a prediction about the causal influence in a new, as yet unobserved dataset.)
    a. Presuming that all other variables are similar to their state in an existing dataset? (This is to predict the setting-specific consequence of change in a causal factor, even though one may not have a real manifestation to test the prediction.)
    b. Under a (possibly) different set of states of the other variables? (This is the prediction of an influence in a future setting where other factors may have changed, needing the ability to sort out the effects of the causative variable in question from other potential influences on the outcome as well as the ability of other states to affect the magnitude of causal influence of the target variable.)

These get progressively more challenging to answer fully and unambiguously as one goes down the list. The list represents progressively more generalizability of the observed patterns (or rather of the explanation proffered for those patterns) to other settings, where the level of the causative agent is different and then where the values of other factors that may enhance, inhibit, or interact are also different from the setting observed. But it is the final question that one wishes to answer in order to judge the impact of alteration of an exposure standard.

In my view, the important final lesson is to avoid conflating the above set of questions. We can develop evidence for some causal role, largely on the grounds of repletion of apparent effects across studies with different conditions, or unambiguous effects at high dose, plus biological understanding of the plausibility of toxicity-generating processes. To judge the present impact of present levels is a step farther, and to project the changes in impacts from possible changes in levels is yet a different and more challenging question. It is not that possible answers cannot be identified, but that they should be interpreted with awareness of how the limits to interpretation of the causation process and its generalizability can lead to uncertainties that need to be factored in to decisions.

<u>Questions from Drs. Packham, Frampton, and Boylan</u>

I have taken my role to be to provide focused comments on particular technical issues, rather than to act as a reviewer of the document as a whole. Many of the questions from these CASAC members are of a more general reviewing nature, and I have not been able to address them. Some questions from Dr. Frampton (Questions 1, 3 and 4) bear on issues similar to those raised by Dr. Cox, and my response above to those is meant to cover the same ground.

Questions from Dr. Lange

I fear that lack of sufficient time adequately to delve into the calculations referred to in Dr. Lange's questions preclude me from addressing them with sufficient rigor at this time.

I will only note that, aside from the purely statistical aspects of the inferences among exposures measured at different durations, there is the toxicological aspect that the dependence of effects on the particular time-course of exposure plays a role, and such dependence is usually treated with asumptions about time-averaging that may be inconsistent with the complexity of underlying dynamics of uptake and clearance as well as damage and repair that ultimately dictate whether and when toxicity may be engendered. That is, the reasons that effects might appear in longer term exposure at levels below those causing impacts from short exposures depend on how the balance of damage and repair processes operate, as well as on whether the longer term risks arise owing to more chances for essentially short-term stochastic events with more chronic consequences once incurred to occur. One needs to ask whether a longer term effect results because of the increased chances that somewhere in the span of averaging time it happened that a set of short-term peaks happened in close enough time-spacing to have some cumulative effect that more widely spaced peaks would not engender, or whether the longer-term effects really depend on inexorable and continuous exposures, even to lower levels. All these questions have to do with averaging time, and any assumptions or calculation methods that entail averaging necessarily invoke some underlying hypothesis about how the effect is attributable to some aspect of the long-term average, despite differences from case to case in the time-varying moment-by-moment exposures. I addressed this general issue in a paper on acute inhalation toxicity (Rhomberg, LR. 2009. "Uptake kinetics, species differences, and the determination of equivalent combinations of air concentration and exposure duration for assessment of acute inhalation toxicity." Hum. Ecol. Risk Assess. 15(6):1099-1145), but the issues explored therein also apply to the comparison of short-term and long-term exposure as causes of effects.