December 17, 2018

John Vandenberg, Ph.D.

Director

National Center for Environmental Assessment

Research Triangle Park Division

Dr. Vandenberg,

Thanks so much again to you and your team for an extremely clear, thorough, and useful presentation and for helpful answers at our recent CASAC PM public meeting. Among other things, I greatly appreciated the clear visualizations that your deck provided for communicating complex statistical information and important ideas in clear and accessible ways.

I would like to follow up as promised on some key definitional issues that we started to discuss but lacked time to thoroughly discuss and resolve. I want to make sure that I understand exactly what the five causal determination categories and the term "independent effect" are intended to mean, and how they are intended to be applied to evidence. For this purpose, I find it helpful to consider simplified thought experiments, where many real-world complexities are stripped away to better highlight the intended meaning and application of these terms in contexts simple enough to make all reasoning explicit. I would greatly appreciate answers to the following hypothetical conceptual questions, which are intended to help achieve greater clarity and shared understanding of exactly what is intended to be communicated when these causal determination categories, or labels, are used to communicate ISA findings.

Questions on definition of "independent effect"

1. In the structural equation model

    Risk = PM*Copollutant

    Copollutant = PM

    where these are causal equations (i.e., structural equations) with the explicit causal interpretation that changing a variable on the right-hand side of an equation causes the variable

on its left-hand side to change to make the equality hold again, how is the independent effect of *PM* on *Risk* defined? For example, if the initial situation is *PM* = 0.5 and *Copollutant* = 0.2 (implying that *Risk* = 0.1), then what would be the independent effect on *Risk* of reducing *PM* from 0.5 to 0.4?

2. In the structural equation model

    > *Risk = Copollutant*
    >
    > *Copollutant = PM*

    (corresponding to the diagram *PM* → *Copollutant* → *Risk*), how is the independent effect of *PM* on *Risk* defined? Is it zero, because changing *PM* alone while holding *Copollutant* fixed would have no impact on risk? Or is it the change in risk caused by changing *PM* and allowing *Copollutant* to respond realistically, i.e., the "total effect" (mediated by change in the co-pollutant) rather than the direct effect? Or does "independent effect" mean something else?

3. In the structural equation model

    > *Risk = Exposure*Poverty*,

    where all three variables are scaled to run from 0 to 1, how is the independent effect of *Exposure* on *Risk* defined? For example, if the initial situation is *Exposure* = 0.5 and *Poverty* = 0.2, what would be the independent effect on *Risk* of reducing *Exposure* from 0.5 to 0.4? (Would it be 0.2*(0.5-0.4) = 0.02, because the level of *Poverty* is 0.2, thus making the "independent effect"of *Exposure*depend on the level of *Poverty*? Would it be 0, because the contribution of a change in *Exposure* alone, in the absence of poverty, would be zero? Or is it something else?)

Questions on definitions of causal determination categories

1. *Causality and preventability.* Does the determination that a PM exposure-response relationship is "Causal" imply that reducing exposure would reduce response?

    a. Conversely, might the "Causal" category be an appropriate designation for a PM concentration-response (C-R) association even if it is known that reducing PM would *not*

change the probability distribution of the response in the exposed population? If so, under what conditions would this be appropriate?

b. More specifically, does a "Causal" determination imply that reducing exposure alone (without changing anything else such as socioeconomic, co-morbidity, co-pollutant, or weather variables) would necessarily reduce any or all of following: prevalence, incidence, average annual frequency per 100,000 capita-years, or age-specific hazard functions for the response in exposed populations?

c. Might the "Causal" category be an appropriate designation for a PM concentration-response (C-R) association even if it is not known whether reducing PM would change the response (or the probability distribution of the response) in the exposed population?

d. *Direct vs. total causation.* Does the determination that a PM exposure-response relationship is causal imply that reducing exposure would reduce response even if other causally relevant factors (e.g., socioeconomic, co-morbidity, co-pollutant, or weather variables) were held fixed at their current values?

e. More generally, what are the empirically testable implications or predictions of a "Causal" designation?

2. *Strength of causal relationships and sizes of effects.* Does the determination that a PM exposure-response relationship is causal imply that reducing exposure by, say, 10 µg/m^3, must reduce response by at least a certain positive amount? In other words, is there any lower limit to how small a change in health effects caused by a given reduction in exposure can be to make "Causal" the appropriate determination? (To use an extreme example, if eliminating exposure completely were to lengthen the life expectancy of just one person by one trillionth of a second, but had no effect on anyone else, would that suffice to designate the C-R relationship as causal for the population? If not, is there a minimum size of effect that must be achieved for the "Causal" label to be appropriate?) Is the five-point categorization intended to convey any information about effect sizes or strength of association?

3. *Homogeneity of causal relationships.* Does the category "Causal relationship" mean the same thing as "Causal relationship for 100% of the members of the exposed population"?

a. Conversely, does the category "Causal relationship" mean the same thing as "Causal relationship for at least one member of the exposed population"?

b. Can an exposure-response or C-R association be causal for part of an exposed population (e.g., men over 70 years old with COPD) without being causal for other parts of the population (e.g., healthy women under 30)?

c. If so, is there a minimum size or fraction of the population for whom the C-R relationship must be causal in order to imply that "Causal" is the correct designation for that relationship in the exposed population as a whole?

d. Might evidence of a C-R relationship be causal for some subpopulations (e.g., COPD patients) but only "Suggestive of, but not sufficient to infer, a causal relationship" for other subpopulations? Why or why not?

e. If different causal determination categories apply to different subpopulations, how should the causal determination category for the population as a whole be determined from the causal determination categories of its subpopulations?

4. *Are causal determination categories mutually exclusive and collectively exhaustive?*

a. Is it possible for evidence to be both "Suggestive of, but not sufficient to infer, a causal relationship" and also "Inadequate to infer the presence of absence of a causal relationship"? Why or why not? (At present, I do not see why both descriptions might not apply simultaneously, or why both might not also be compatible with the "likely to be causal" category.)

b. More generally, what prevents a body of evidence from being correctly described by more than one of these categories? The descriptions given in the framework for causality determinations described in the Preamble to the ISAs (U.S. EPA, 2015) and in Table P-2 appear to allow for considerable overlap between some of the five categories. For example, the "Causal" category includes as examples "observational studies that cannot be explained by plausible alternatives or that are supported by other lines of evidence (e.g., animal studies or mode of action information). Generally, the determination is based on multiple high-quality studies conducted by multiple research groups." The "Likely to be causal" category includes this: "animal toxicological evidence from multiple studies from different laboratories demonstrate effects, but limited or no human data are available. Generally, the determination is based on multiple high-quality studies." Now, suppose that a body of evidence consists of

i. Observational studies in humans that can plausibly be explained by plausible alternatives such as an unmeasured confounder or coincident historical trends; and

ii. Supporting animal toxicological evidence from multiple studies from different laboratories that demonstrate species-specific effects based on multiple high-quality studies conducted by multiple research groups.

Which category applies in this case? On the one hand, the evidence satisfies the description "observational studies that… are supported by other lines of evidence (e.g., animal studies or mode of action information). Generally, the determination is based on multiple high-quality studies conducted by multiple research groups." That would indicate that it belongs to the "Causal" category. On the other hand, it also satisfies the description "animal toxicological evidence from multiple studies from different laboratories demonstrate effects, but limited… human data are available. Generally, the determination is based on multiple high-quality studies." Thus, it seems it should also belong to the "Likely to be causal" category. To which category should such a body of evidence that is described by more than one be assigned, and on what basis?

c. In the example just given, the body of evidence consisted of observational data in humans that can plausibly be explained by alternatives such as an unmeasured confounder or coincident historical trends, together with animal data showing a species-specific response. Shouldn't this be categorized as "Inadequate to infer a causal relationship" rather than (or in addition to) being categorized as "Causal" and/or "Likely to be causal" as Table P-2 seems to require?

d. Two of the causal determination categories are "Likely to be a causal relationship" and "Not likely to be a causal relationship." Why doesn't at least one of these labels apply to each body of evidence?

e. The description for "Not likely to be a causal relationship" says "Evidence indicates there is no causal relationship with relevant pollutant exposures." What evidence would indicate that there is no causal relationship (rather than that there is no detected causal relationship)? This seems to require proving a negative.

f. Suppose that an initial body of evidence consists of animal toxicological evidence from a relatively few (but multiple) high-quality studies from different laboratories that demonstrate effects, but that no human data are available. This matches one of the

cases described as "Likely to be a causal relationship" in Table P-2. Now suppose that two further supporting studies are added: a high-quality epidemiologic study that shows an association with a given health outcome; and a high-quality toxicological study that shows effects relevant to humans in an animal species. The evidence now matches one of the cases described as "Suggestive of, but not sufficient to infer, a causal relationship" in Table P-2. Should the addition of these two new supportive studies result in a downgrade of the evidence from its previous label of "Likely to be a causal relationship" to a new label of "Suggestive of, but not sufficient to infer, a causal relationship," to match the classification of case descriptions in Table P-2? Why or why not?

5. *Operational definitions of "adequacy" and "sufficiency".*

   a. What are the defining (operationally testable and independently verifiable) conditions that make evidence "sufficient to infer a causal relationship"? What is the operational definition of this category?

   b. Are there also defining conditions that make evidence "insufficient to infer a causal relationship?" If so, what are they?

   c. What are the defining conditions that make evidence "Inadequate to infer the presence or absence of a causal relationship"? What is the operational definition of this category?

   d. Are there also defining conditions that make evidence sufficient to infer absence of a causal relationship? If so, what are they?

6. *Certainty of causal relationship category.* Does the category "Causal relationship" mean the same thing as "Causal relationship with 100% certainty, probability, or confidence"? If not, is there a threshold for certainty, probability, or confidence below which it would be inappropriate to call a relationship "causal"?

7. *Categorizing simple cases where all relevant information is known.* Consider an example in which it is known that that an observed C-R relationship is either causal (if there is no unmeasured confounder that explains it) or not (otherwise). There are no other relevant facts, considerations, or lines of evidence. The probability of such an unmeasured confounder has been bounded by data analysis of multiple past studies as being no greater than p, where p is a number between 0 and 1.

a. For what values of p should the C-R relationship be categorized as "Causal"? Is there a smallest value of p (the probability that the relationship is not causal) that is required for the "Causal" label to be applicable?

b. Similarly, suppose that different data analyses establish that the probability of an unmeasured confounder is no less than q, where q is a number between 0 and 1. For what values of q should the relationship *not* be considered causal?

c. Are there values of p and q for which the causal determination category is ambiguous?

d. If further research determines that the probability of an unmeasured confounder is in fact r (to two decimal places), where r is a number between 0 and 1, then for what values of r is each of the five causal determination categories the correct description? For example, if $r = 0.5$, which causal determination category would be the correct one to use, and why, assuming that there are no other relevant uncertainties or facts?

8. *Quantity of evidence needed for a causal determination.* Suppose that each of 10 independent studies (possibly including diverse types of evidence, e.g., epidemiological, toxicological, and clinical studies) concludes that the hypothesis of no causal relationship between C and R can be rejected with at least 95% statistical confidence. For simplicity, assume that this is the totality of the available evidence. (Thus, no studies have reached a different conclusion.) Would this constitute sufficient evidence to conclude that the C-R relationship should be classified as causal? Would 2 such studies be enough? In this simple setting, is there a minimum number of such studies that would benecessary and sufficient to warrant labeling the studied C-R relationship as "Causal" even though 100% certainty can never be achieved?

9. *Discordant evidence.* Suppose that 7 studies estimate a significant positive C-R relationship at the 95% confidence level (e.g., the 95% confidence interval for the relative risk is entirely to the right of 1), but another 3 studies estimate significant negative C-R relationships (95% confidence intervals entirely to the left of 1). Upon close scrutiny, all studies appear to have the same high quality and their conclusions appear to be equally sound and credible. If this were the only relevant evidence, then what conclusions about causal determination category, if any, should be drawn from such discordant evidence? If the numbers were changed (e.g., to 1000 studies reaching one conclusion and 2 reaching the opposite conclusion), how, if at all, should the resulting causal determination category change in response?

10. *Updating evidence categorizations.* Is it possible that a C-R relationship that is presently classified as causal might later be reclassified in light of additional evidence? Are there any

restrictions on how likely this possibility must be in order for the "Causal" classification to be applied? For example, is a current designation of "Causal" for a relationship inconsistent with a judgment that there is a 90% probability that the relationship will be reclassified as "Inadequate to infer the presence of absence of a causal relationship" as soon as an accountability study now underway is concluded? What restrictions, if any, does a current designation of "Causal" imply for possibilities and probabilities of future reclassifications?

I greatly appreciate any answers you can provide. If the answers (or questions) are unclear for any of these, or if some of the questions are simply not appropriate for understanding the intended use of the causal determination categories and the term "independent effect," I would also appreciate knowing that. I am hopeful that discussing the answers will clarify the intended meanings of these terms as they are used to interpret and communicate the scientific information in the ISA to policy makers and other recipients.

Dr. Tony Cox
Chair
Clean Air Scientific Advisory Committee