

**To:** Professor Lianne Sheppard

**From:** Tony Cox

**Re:** References on Vetting the Causal Determination Framework in the Draft PM PA and ISA

**Date:** 10-23-2019

Thank you so much for your reference to the IOM report (Institute of Medicine 2008. Improving the Presumptive Disability Decision-Making Process for Veterans, Chapter 8. Washington, DC: The National Academies Press. <https://doi.org/10.17226/11908>). That reference *proposes* an approach, and compares it to earlier proposals (see section on “Comparison of the Committee’s Proposed and Previous Frameworks”), but I find nothing about empirical *validation* of the proposed approach (e.g., false positive and false negative rates, when applied to real data, or perhaps results similar to Figure 1 of Dorie et al. 2018, <https://arxiv.org/pdf/1707.02641.pdf>). I am seeking empirical evidence on how well the proposed approach works in practice, especially for air pollution health effects.

It appears to me that Chapter 8 presents a proposed framework based largely on the Bradford Hill considerations (addressing attributive causation, but not manipulative or interventional causation) and on assumptions that most experts in causality reject (e.g., that “Association, especially association adjusted for potential confounders, is evidence for the causal claim” (p. 193); this has been widely recognized as invalid since at least the early 1960s, when threats to internal validity of causal conclusions based on associations (such as coincident historical trends, regression to the mean, etc.) were enumerated and discussed as testable non-causal hypotheses by Campbell and Stanley in [\*Experimental and Quasi-Experimental Designs for Research\*](#)). To my knowledge, the proposed WoE framework has not been used by any successful team in any causal inference competition. I have heard and read many times that it is “well vetted,” but now I would appreciate very specific guidance on exactly where its performance is rigorously evaluated and what the results are – e.g., is its balanced accuracy more or less than 50% for “causal” determinations?

I very much appreciate your help and value your professional integrity, as expressed in our call today. As you know, I have struggled to understand the basis for the repeated claim that “This [WoE] framework is an appropriate tool for drawing causal conclusions.” I also thirst for details backing up the claim that “it has been well-vetted over more than a decade.” To me, “well-vetted” means that it has been shown empirically, as well as in principle, to have good performance e.g., as measured by low rates of type and type 2 error (or, more generally, by confusion matrices) for its determinations. I look forward to your help in identifying thorough evaluations of the theoretical and empirical performance of the WoE framework now being used in the current Draft ISA and PA.