Published in final edited form as:

Stat Med. 2013 November 10; 32(25): 4334–4337.

Additional thoughts on causal inference, probability theory, and graphical insights

Stuart G. Baker*

*Biometry Research Group, National Cancer Institute, Bethesda, MD 20892 USA. sb16i@nih.gov

I was delighted that Judea Pearl wrote a commentary [1] on my paper “Causal inference, probability theory, and graphical insights”[2] because it provides a springboard for improving clarity. Judea Pearl and I tackle the problem of causal inference from different research backgrounds which may explain the difference in our viewpoints.

Pearl wrote that the paired availability design “involves counterfactuals and other non-probabilistic notions,” explaining the latter as “relations or parameters that cannot be defined in terms of observed variables,” and adding “The restriction to observed variables is important for, otherwise, everything would become probabilistic” I find these comments puzzling because the paired availability design involves computing estimated probabilities of unobserved events related to counterfactuals based on estimated probabilities of observed events. The PAD-Plot diagrams the probabilities of observed events that sum over probabilities of unobserved events associated with counterfactual variables. In interpreting the PAD-Plot, one can simply replace the population probabilities of observed events with their obvious estimates (that are maximum likelihood for a perfect fit of the saturated model). Thus the PAD-plot shows graphically the computation of estimated probabilities of unobserved events related to counterfactuals (most notably the estimated treatment effect in the relevant principal stratum) as a function of estimated probabilities of observed events. In this sense, the paired availability design involves a “restriction” to observed variables, and probabilities of events related to counterfactual are defined in terms of unobserved variables.

Pearl noted that an instrumental variable cannot be defined in terms of a joint distribution of observed variables. I agree. Pearl also noted that causal graph (*e*) in Figure 1 augmented with an arrow from *Z* to *Y* can generate a probability distribution in which *Z* is an instrument. I agree in the sense that causal graph (*d*) is a special case of Pearl's augmented causal graph. However, I am unclear as to the relevance of these statements to my discussion. I define an instrumental variable based on assumptions summarized in causal graph (*d*) about a joint probability distribution of observed and unobserved variables. My causal graph for an instrumental variable is identical to the causal graph that Pearl used to discuss instrumental variables in the context of bias amplification.[3]

Pearl agreed with me that d-separation (a useful and clever method) is a probabilistic tool. However, I was puzzled when Pearl added that d-separation has “nothing to do with causation nor ‘bias.’” My understanding is that d-separation indicates the relevant variables for adjustment. Adjusting for these relevant variables yields an unbiased estimate of treatment effect in the sense of having the same expectation as an estimate of treatment effect obtained from a randomized trial. The BK-plot illustrates how adjusting for a single

relevant variable leads to an unbiased estimate. I was also puzzled when Pearl wrote “Appendix A does not provide a proof that adjustment on U is not appropriate.” Appendix A proves that adjusting on U makes X and Y dependent on the back door path, which means the adjustment on U is inappropriate because it biases the estimated effect of X and Y on the relevant front door path.

Pearl was surprised by my claim that the paired availability design [4] does not fit into the causal graph framework. Pearl correctly noted that the paired availability design involves principal stratification.[5] Pearl also wrote that principal stratification is “a counterfactual framework that fit perfectly and, in fact, is subsumed by the causal graph framework. A structural causal model represents all counterfactuals that may possibly be defined among the variables in the model and, therefore subsumes any design based on these counterfactuals.” However the ability to represent all counterfactuals is not sufficient for estimating treatment effect under the paired availability design. The key to the paired availability design is the appropriateness of the assumptions. Recall that the principal strata are n (never-receivers) who would receive treatment T_0 if in either time period, a (always-receivers) who would receive treatment T_1 if in either time period, c (consistent-receivers) who would receive T_0 (T_1) if in the time period when T_1 is less (more) available, and i (inconsistent-receivers) who would receive T_1 (T_0) if in the time period when T_1 is less (more) available. The paired availability design requires the following two assumptions involving principal strata: (1) the probability of outcome does not change over time periods among participants in principal strata n and a and (2) under fixed availability there are no participants in principal stratum i , and under random availability the probability of outcome is the same under principal strata c and i . Also the paired availability design requires additional assumptions (stable population, stable ancillary care, stable disease progression, stable evaluation) to justify analyzing data from before-and-after time periods as if they were randomization groups. [4] How do causal graphs encode these assumptions?

Pearl reported that a search in Google Scholar listed only a handful of entries on the paired availability design. Because the paired availability design may be unfamiliar to many readers, I provide the following brief background. The paired availability design is a method for estimating the effect of treatment (T_1 versus T_0) from historical controls among multiple medical centers. It avoids the self-selection bias of a traditional analysis of historical controls which compares outcomes among persons receiving T_1 in the second time period with outcomes among persons receiving T_0 in the first time period. Instead the paired availability design compares outcomes among *all* eligible persons in two time periods and uses principal stratification with plausible assumptions to estimate the effect of receipt of T_1 instead of T_0 . (Permutt and Hebel [6], Angrist, Imbens, and Rubin [7] and Cuzick, Edwards, and Segnan independently [8] developed related principal stratification models for all-or-none compliance in two arms of a randomized trial). Researchers should consider the paired availability design when a randomized trial cannot be conducted and the outcome is observed in a short time period (to minimize changes in ancillary care over time). To reduce the possibility for bias, investigators should choose medical centers that have little in- or out-migration such as geographically isolated hospitals and army medical centers. Building on earlier work involving principal stratification with all-or-none compliance [9], in 1994

Karen Lindeman and I proposed the paired availability design to estimate the effect of epidural analgesia on the rate of caesarean section.[10] By 2001, there were a sufficient number of before-and-after studies with the requisite data for us to conduct an informative analysis using the paired availability design.[11,12] In 2013 we updated the analysis with more recent data and methodology.[4] The paired availability design gave an estimate of the effect of epidural analgesia on the rate of caesarean section that was similar to the estimate from a meta-analysis of randomized trials (near zero with narrow confidence intervals) but different from the estimates from two multivariate observational studies with concurrent controls (greater than zero with confidence intervals that excluded zero). We think the estimates from the multivariate observational studies were biased because they did not adjust for an unmeasured confounder, namely intense pain.[4] I have also proposed using the paired availability design to evaluate cancer screening when the short-term outcome is cancers arising in the interval between screens. [13] However the paired availability design is not well known in other areas of medicine. The *Encyclopedia of Biostatistics* mentions the paired availability design under the heading of “Nonrandomized trials,”[14] and the reference book *Medical Statistics from A to Z. A Guide for Clinicians and Medical Students* includes an entry for the paired availability design. [15]

In summarizing his viewpoint, Pearl stated “Modern casual inference owes much of its progress to a strict and crisp distinction between probabilistic and causal information. This distinction recognizes that probability theory is insufficient for posing causal questions, let alone answering them and dictates that every exercise in causal inference must commence with some extra knowledge that cannot be expressed in probability alone.” I find probability theory with subject matter assumptions to be a useful framework for obtaining causal estimates of treatment effect in both multivariate observational studies with concurrent controls and the paired availability design with historical controls. The former involves subject-matter assumptions about probabilistic relationships among candidate variables for adjustment. The latter involves subject-matter assumptions related to probabilities of outcome given principal strata and time period and also probabilities of being in various principal strata. However there are many types of scientific inquiry that require a very different type of thinking about causality. One example of particular interest to me is whether the proximal cause of cancer (for example after exposure to a carcinogen) is successive mutations of somatic cells (the somatic mutation theory) or the disruption of the intercellular communication needed for tissue organization (the tissue organization field theory).[16-18] This scientific inquiry involves a choice between thinking about causation at the level of the cell or the level of the tissue. From the tissue-level perspective a famous analogy is “cancer is no more a disease of cells than a traffic jam is a disease of cars. A lifetime of study of the internal-combustion engine would not help anyone to understand our traffic problems.”[19] Another important issue is how well each theory of carcinogenesis explains various experimental results, such as the development of tumors when small-pore filters, but not large-pore filters, are inserted under the skin of mice.[20] This is a very different type of thinking about causality than found in statistics journals.

On the topic of the BK-Plot, I would like to report on some relevant publications that came to my attention after the publication of my paper “Causal inference, probability theory, and

graphical insights.” By way of background, after I had formulated the BK-Plot I discovered that JW Jeon [21] and colleagues had published the same plot earlier in 1987 in the *Journal of the Korean Statistical Society*, which I noted in my original publication. I have now learned that Arjan Tan [22] published the same plot in 1986 in *The College Mathematics Journal*; it was based on a geometric interpretation of the weighted mean developed by Larry Hoehn. [23] In a 2001 article, Larry Lesser [24] discussed various graphical methods illustrating the reversal with Simpson's paradox. In 2013 Kady Schneider and Jürgen Symanzik [25] published an applet illustrating the BK-Plot.

Acknowledgements

This work was supported by the National Institutes of Health.

References

1. Pearl J. Comment on “Causal inference, probability theory, and graphical insights” by Stuart G. Baker. *Statistics in Medicine*. 2013 Forthcoming.
2. Baker SG. Causal inference, probability theory, and graphical insights. *Statistics in Medicine*. 2013 early view. doi: 10.1002.
3. Pearl, J. In Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010). Association for Uncertainty in Artificial Intelligence; Corvallis: 2010. On a class of bias-amplifying variables that endanger effect estimates.; p. 425-432.
4. Baker SG, Lindeman KL. Revisiting a discrepant result: a propensity score analysis, the paired availability design for historical controls, and a meta-analysis of randomized trials. *Journal of Causal Inference*. 2013; 1:51–82.
5. Frangakis CE, Rubin DB. Principle stratification in causal inference. *Biometrics*. 2002;58:21–29. [PubMed: 11890327]
6. Permutt T, Hebel R. Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics*. 1989; 45:619–622. [PubMed: 2669989]
7. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. 1996; 92:444–455.
8. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Statistics in Medicine*. 1997; 16:1017–1029. [PubMed: 9160496]
9. Baker, SG. A new way of analyzing randomized consent designs. Department of Biostatistics, Harvard School of Public Health; 1983. Unpublished manuscript A copy is available upon request
10. Baker SG, Lindeman KS. The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine*. 1994; 13:2269–2278. [PubMed: 7846425]
11. Baker SG, Lindeman KS. Rethinking historical controls. *Biostatistics*. 2001; 2:383–396. [PubMed: 12933631]
12. Baker SG, Kramer BS, Lindeman KS. The paired availability design: If you can't randomize, perhaps this applies. *Chance*. 2006; 19:57–60.
13. Baker SG. Improving the biomarker pipeline to develop and evaluate cancer screening tests. *Journal of the National Cancer Institute*. 2009; 101:1116–1119. [PubMed: 19574417]
14. Gehan, EA. Encyclopedia of Biostatistics. Wiley Online Library; 2005. Nonrandomized trials..
15. Everitt, BS. Medical Statistics from A to Z. A Guide for Clinicians and Medical Students. Cambridge University Press; Cambridge: 2006.
16. Baker SG. Paradoxes in carcinogenesis should spur new avenues of research: An historical perspective. *Disruptive Science and Technology*. 2012; 1:100–107.
17. Baker SG. Paradox-driven cancer research. *Disruptive Science and Technology*. 2013; 1:143–148.
18. Soto AM, Sonnenschein CS. The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. *Bioessays*. 2011; 33:332–340. [PubMed: 21503935]

19. Smithers DW. Cancer. An attack on cytologism. *The Lancet*. 1962; 279:493–499.
20. Karp RD, Johnson KH, Buoen LC, et al. Tumorigenesis by Millipore filters in mice: histology and ultrastructure of tissue reactions as related to pore size. *Journal of the National Cancer Institute*. 1973; 51:1275–1285. [PubMed: 4583375]
21. Jeon JW, Chung HY, Bae JS. Chances of Simpson's paradox. *Journal of the Korean Statistical Society*. 1987; 16:117–125.
22. Tan A. A geometric interpretation of Simpson's paradox. *The College Mathematics Journal*. 1986; 17:340–341.
23. Hoehn L. A geometrical interpretation of the weighted mean. *The College Mathematics Journal*. 1984; 15:135–139.
24. Lesser, L. Representations of reversal: An exploration of Simpson's paradox.. In: Cuoco, Albert A.; Curcio, Frances R., editors. *The Roles of Representation in School Mathematics*. National Council of Teachers of Mathematics; Reston, VA: 2001. p. 129-145.
25. Schneider K, Symanzik J. An applet for the investigation of Simpson's paradox. *Journal of Statistics Education*. 2013; 21:1.