

Comments and Controversies: Graphical models, potential outcomes and causal inference: Comment on Lindquist and Sobel

Judea Pearl
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
judea@cs.ucla.edu

June 28, 2011

Dear Editor,

I read with interest the comment by Lindquist and Sobel (L&S, 2011) entitled: “Graphical models, potential outcomes and causal inference” (*NeuroImage*, online 2010) in which they advocate the use of counterfactual language to explicate causal assumptions, and raise doubts on whether graphical models are generally useful for estimating causal effects. Their comment creates the impression, perhaps unintentionally, that counterfactual language is somehow superior, more rigorous or more principled than the graphical language used by structural equation modelers (SEM) in fMRI research. The purpose of this communication is to correct any such impression and to supplement L&S comment with proven mathematical results regarding the relations between the two notational systems.

It has been proven (Balke and Pearl, 1994; Galles and Pearl, 1998; Halpern, 1998; Pearl, 2009, Ch.7) that the two notational systems are logically equivalent in the sense that a theorem in one is a theorem in the other, and an assumption in one has a parallel interpretation in the other. The translation between the two is given by two simple rules (Pearl, 2009, p. 101) that rewrite assumptions conveyed in graphical form into symbolic counterfactual notation. In particular, assumptions A1–A4(b) that L&S present in their paper are faithfully represented by the causal chain $Z \rightarrow X \rightarrow Y$ which L&S aim to

replace or discredit.

To facilitate readers examination, I copy these assumptions below

Assumption 1: the existence of the potential outcomes $x(z)$ and $Y(z, x)$ for all z and x ,

Assumption 2: $Y(0, x) = Y(1, x)$ for all x , expressing the idea that Z does not directly cause Y ,

Assumption 3: $X = X(Z), Y(z) = Y(z, X(z)), Y = Y(Z, X(Z))$,

Assumption 4a: $Y(z, x), X(z) \perp\!\!\!\perp Z$ for all z, x

Assumption 4b: $Y(z, x) \perp\!\!\!\perp X|Z$ for all z, x .

These assumptions can be derived from the counterfactual reading of causal chain $Z \rightarrow X \rightarrow Y$ which simply specifies what factors participate in determining the value of each variable in the model and whether omitted factors are dependent of each other. In our example, the graph specifies that:

1. Y is determined by X only,
2. X is determined by Z only, and
3. All functional relationships are further modified by omitted factors (not shown explicitly in the graph) that are assumed to be mutually independent yet arbitrarily distributed.

No additional assumptions beyond these three are needed to derive *all* the causal and counterfactual conclusions obtained by L&S, and all subsequent causal and counterfactual relations estimable from data with the help of assumptions A1–A4(b) (see Pearl, 2010, pp. 126–127 for explicit derivation).

It would be instructive for fMRI researchers to examine the counterfactual assumptions A1–A4(b) above, compare them to their graphical encoding in the causal chain $Z \rightarrow X \rightarrow Y$, and assess which notational representation is more transparent, rigorous, explicit, and conducive to meaningful scientific discourse.

While if it is true, as noted by L&S, that some SEM researchers have confused the causal reading of SEM with regressional interpretations of the parameters (see Pearl, 2009, pp. 135–138 for the history and reason of this confusion), the fact remains that, in their correct SEM interpretation, DAGs offer a parsimonious notational system capable of encoding vividly the very same counterfactual assumptions that L&S consider essential for causal inference.

This general translation from DAG's to counterfactuals gives researchers the options of either explicating causal assumptions algebraically, in the parenthetical notation advocates by L&S, or derive these assumptions from the DAG when the need arises, or, assuming they wish to stay close to scientific intuition, keep them in graphical language and utilize the inferential machinery that this language provides. In view of the many results that this machinery has spawned in the past two decades (Greenland et al., 1999; Pearl, 1995, 2009; Spirtes et al., 2000, see also Pearl, 2010 for a recent survey), fMRI researchers should be encouraged to continue using their familiar SEM language and be assured that the results thus obtained are no less valid than those derived in the counterfactual language. They are likely in fact to be more valid, considering the opaque mathematical form of the latter, as exemplified by assumptions A1–A4(b) above, and the transparency of their graphical counterparts.

References

- BALKE, A. and PEARL, J. (1994). Counterfactual probabilities: Computational methods, bounds, and applications. In *Uncertainty in Artificial Intelligence 10* (R. L. de Mantaras and D. Poole, eds.). Morgan Kaufmann, San Mateo, CA, 46–54.
- GALLES, D. and PEARL, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundation of Science* **3** 151–182.
- GREENLAND, S., PEARL, J. and ROBINS, J. (1999). Causal diagrams for epidemiologic research. *Epidemiology* **10** 37–48.
- HALPERN, J. (1998). Axiomatizing causal reasoning. In *Uncertainty in Artificial Intelligence* (G. Cooper and S. Moral, eds.). Morgan Kaufmann, San Francisco, CA, 202–210. Also, *Journal of Artificial Intelligence Research* 12:3, 17–37, 2000.
- LINDQUIST, M. A. and SOBEL, M. E. (2011). Graphical models, potential outcomes and causal inference: Comment on Ramsey, Spirtes and Glymour. *NeuroImage* **57** 334–336, DOI: 10.1016/j.neuroimage.2010.10.020 (online 2010).
- PEARL, J. (1995). Causal diagrams for empirical research. *Biometrika* **82** 669–710.
- PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

PEARL, J. (2010). The foundations of causal inference. *Sociological Methodology* **40** 75–149.

SPIRITES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*. 2nd ed. MIT Press, Cambridge, MA.