

Published in *Biometrika*,  
82, 695-698, 1995

**Comment on Judea Pearl's Paper,  
"Causal Diagrams for Empirical Research"**

James M. Robins

Departments of Epidemiology and Biostatistics

Harvard School of Public Health

Boston, MA 02115 USA

I wish to thank the editor, Phil Dawid, for his invitation to discuss Judea Pearl's seminal paper. Professor Pearl carried out two tasks. In the first, in Sec. 2, using some results of Spirtes et al. (1993), he showed that a non-parametric structural equations model depicted as DAG  $G$  implies the causal effect of any subset  $X \subset G$  on  $Y \subset G$  is a functional of (i) the distribution function  $P_G$  of the variables in  $G$  and (ii) the partial ordering of these variables induced by the directed graph. This functional is the g-computation algorithm functional, hereafter g-functional, of Robins (1986, p. 423). In the second, in Secs. 3-5, only a subset of the variables in  $G$  is observed. Given known conditional independence restrictions on  $P_G$  encoded as missing arrows on  $G$ , Pearl developed elegant graphical inference rules for determining identifiability of the g-functional from the law of the observed subset. Task 2 requires no reference to structural models or to causality. A potential problem with Pearl's formulation is that his structural model unsuitably implies that all variables in  $G$ , including concomitants like age or gender, are potentially manipulable. In Section 1.1 below, I describe a less restrictive model that avoids this problem but, when true, still implies that the g-functional equals the effect of the treatments  $X$  of interest on  $Y$ . This critique of Pearl's structural model is unconnected with his graphical inference rules, which were his main focus and are remarkable and path-breaking, going far beyond my own and others' results (Robins, 1986, Sec. 8 and Appendix F).

## **1. Task 1**

In Robins (1986, 1987), I proposed a set of counterfactual causal models based on event trees, called causally interpreted structured tree graphs, hereafter causal

graphs, that includes Pearl’s non-parametric structural equations model as a special case. These models extended Rubin’s (1978) ”time-independent treatment” model to studies with direct and indirect effects and time-varying treatments, concomitants, and outcomes. In this Section, I describe some of these models.

### 1.1. A Causal Model

Let  $V_i = \{V_{1i}, \dots, V_{Mi}\}$  denote a set of temporally-ordered discrete random variables observed on the  $i^{\text{th}}$  study subject,  $i = 1, \dots, n$ . Let  $X_i = \{X_{1i}, \dots, X_{Ki}\} \subset V_i$  be temporally-ordered, potentially manipulable, treatment variables of interest. The effect of  $X_i$  on an outcome  $Y_i \subset V_i \setminus X_i$  is defined to be  $pr \{Y_i(x) = y\}$  where the counterfactual random variable  $Y_i(x)$  denotes a subject’s  $Y$  value had *all*  $n$  subjects followed the *generalized* treatment regime  $g = x \equiv \{x_1, \dots, x_K\}$ . In Robins (1986), I wrote  $pr \{Y_i(x) = y\}$  as  $pr(y | g = x)$ . Pearl substitutes  $pr(y | \check{x})$ . We regard the  $\{V_i, Y_i(x); x \in \text{support of } X_i\}$ ,  $i = 1, \dots, n$ , as independent and identically distributed copies of random variables  $\{V, Y(x); x \in \text{support of } X\}$  and henceforth suppress the  $i$  subscript.

This formal set-up can accommodate a superpopulation model with deterministic outcomes and counterfactuals as in Rubin (1978): Suppose we regard the  $n$  study subjects as randomly sampled without replacement from a large superpopulation of  $N$  subjects, and our interest is in the causal effect of  $X$  on  $Y$  in the superpopulation. Then, even if for each superpopulation member,  $V$  and  $Y(x)$  are deterministic non-random quantities, nonetheless, in the limit as  $N \rightarrow \infty$  and  $n/N \rightarrow 0$ , we can model the data on the  $n$  study subjects as independent and identically distributed draws from the empirical distribution of the superpopulation.

We now show that  $pr(y | g = x)$  is identified from the law of  $V$  if each component  $X_k$  of  $X$  is assigned at random given the past. Let  $L_k$  be the variables occurring between  $X_{k-1}$  and  $X_k$ , with  $L_1$  being the variables preceding  $X_1$ . Write  $\bar{L}_k = (L_1, \dots, L_k)$ ,  $L = \bar{L}_K$ , and  $\bar{X}_k = (X_1, \dots, X_k)$ . For notational convenience, define  $\bar{X}_0$ ,  $\bar{L}_0$ , and  $\bar{V}_0$  to be variables that are identically 0. In considering task 1 in Robins (1987, Theorem AD.1 and its corollary), I proved the following.

**Theorem:** If, in Dawid’s (1979) conditional independence notation, for all  $k, x$ ,

$$Y(x) \perp\!\!\!\perp X_k \mid \bar{L}_k, \bar{X}_{k-1} = \bar{x}_{k-1} \quad (1.1)$$

$$X = x \Rightarrow Y(x) = Y \quad (1.2)$$

$$pr [X_k = x_k \mid \bar{X}_{k-1} = \bar{x}_{k-1}, \bar{L}_k] \neq 0, \quad (1.3)$$

then

$$pr (y \mid g = x) = h (y \mid g = x) \text{ where} \quad (1.4)$$

$$h (y \mid g = x) \equiv \sum_{\bar{\ell}_K} pr (y \mid \bar{\ell}_K, \bar{x}_K) \prod_{k=1}^K pr (\ell_k \mid \bar{\ell}_{k-1}, \bar{x}_{k-1})$$

is the g-functional for  $x$  on  $y$  based on covariates  $L$ . If  $X$  is univariate,  $h (y \mid g = x) = \sum pr (y \mid x, \ell_1) pr (\ell_1)$  (Rosenbaum and Rubin, 1983).

As in Robins (1987, p.327), I shall refer to  $V$  as a  $R(Y, g = x)$  causal graph whenever Eqs. (1.1)-(1.2) hold, where  $R(Y, g = x)$  stands for "randomized with respect to  $Y$  for treatment  $g = x$  given covariates  $L$ ." In Robins et al. (1992), Eq. (1.1) is called the assumption of no unmeasured confounders given  $L$ . Under the aforementioned superpopulation model, Eq. (1.1) will hold, as  $N \rightarrow \infty$ , in a true sequential randomized trial with  $X$  randomized and  $X_k$ -specific randomization probabilities that depend only on the past  $(\bar{L}_k, \bar{X}_{k-1})$ . In observational studies, (1.1) is untestable; investigators can at best hope to identify covariates  $L$  so that (1.1) is approximately true. Eq. (1.2) is Rubin's (1978) stable unit treatment value assumption. It says  $Y$  and  $Y(x)$  are equal for subjects with  $X = x$ , irrespective of other subjects'  $X$  values. In Robins (1993), I show that

$$h (y \mid g = x) = E \left[ I (X = x) I (Y = y) / \prod_{k=1}^K pr (x_k \mid \bar{x}_{k-1}, \bar{L}_k) \right],$$

whose denominator clarifies the need for (1.3). See also Rosenbaum and Rubin (1983).

## 1.2. Relationship with Pearl's Work

Suppose we represent our ordered variables  $V = \{V_1, \dots, V_M\}$  by a DAG  $G$  that has no missing arrows, so that  $\bar{V}_{m-1} \equiv (V_1, \dots, V_{m-1})$  are  $V_m$ 's parents. Then Pearl's non-parametric structural equation model becomes

$$V_m = f_m (\bar{V}_{m-1}, \epsilon_m), f_m (\cdot, \cdot) \text{ unrestricted, } m = 1, \dots, M \quad (1.5)$$

and

$$\epsilon_m, 1 \leq m \leq M, \text{ are jointly independent} \quad (1.6)$$

Pearl's assumption of missing arrows on  $G$  is (i) more restrictive than (1.5) and (ii) only relevant when faced with unobserved variables as in task 2. We now establish the equivalence between model (1.5)-(1.6) and a particular causal graph, the finest fully randomized causal graph. For any  $X \subset V, x \in \text{support } X$ , let the counterfactual random variable  $V_m(x)$  denote the value of  $V_m$  had  $X$  been manipulated to  $x$ .

**Definitions - Robins (1986, pp. 1419-1423):**(a):  $V$  is a finest causal graph if (i) all one-step ahead counterfactuals  $V_m(\bar{v}_{m-1})$  exist and (ii)  $V$  and the counterfactuals  $V_m(x)$  for any  $X \subset V$  are obtained by recursive substitution from the  $V_m(\bar{v}_{m-1})$ ; e.g.,  $V_3 \equiv V_3\{V_1, V_2(V_1)\}$  and  $V_3(v_1) = V_3\{v_1, V_2(v_1)\}$ . (b): A finest causal graph  $V$  is a finest fully randomized causal graph if for all  $m$ ,

$$\left\{V_{m+1}(\bar{V}_{m-1}, v_m), \dots, V_M(\bar{V}_{m-1}, v_m, \dots, v_{M-1})\right\} \prod V_m | \bar{V}_{m-1} \quad (1.7)$$

For  $V$  to be a finest causal graph, all variables  $V_m \in V$  must be manipulable. Eq. (1.7) essentially says that each  $V_m$  was assigned at random given the past  $\bar{V}_{m-1}$ . In particular, Eq. (1.7) would hold in a true sequential randomized trial in which all variables in  $V$ , not just the treatments  $X$  of interest, are randomly assigned given the past.

**Lemma 1:** (i) Eq. (1.5) is equivalent to  $V$  being a finest causal graph, and (ii) Eqs. (1.5) and (1.6) are jointly equivalent to  $V$  being a finest fully randomized causal graph.

**Proof of Lemma (1i):** If (1.5) holds, define  $V_m(\bar{v}_{m-1})$  to be  $f_m(\bar{v}_{m-1}, \epsilon_m)$ . Conversely, given  $V_m(\bar{v}_{m-1})$ , define  $\epsilon_m = \left\{V_m(\bar{v}_{m-1}); \bar{v}_{m-1} \in \text{support of } \bar{V}_{m-1}\right\}$  and set  $f_m(\bar{v}_{m-1}, \epsilon_m) = V_m(\bar{v}_{m-1})$ . Lemma (1ii) follows by some probability calculations.

" $V$  a finest fully randomized causal graph" implies that  $V$  is a  $R(Y, g = x)$  causal graph, and thus, given (1.3), that  $pr(y | g = x) = h(y | g = x)$ . The converse is false. For example, " $V$  a  $R(Y, g = x)$  causal graph" only requires that the treatments  $X$  of interest must be manipulable.

## 2. Task 2:

Given (1.1)-(1.3), to obtain  $pr(y | g = x)$ , we must compute  $h(y | g = x)$ . However, data often cannot be collected on a subset of the covariates  $L \subset V$  believed sufficient to make (1.1) approximately true. Given a set of *correct* conditional independence restrictions on the law of  $V$ , encoded as missing arrows on a DAG  $G$  over  $V$ , Pearl provides graphical inference rules for determining whether  $h(y | g = x)$  is identified from the observed data. Pearl's graphical inference rules are correct without reference to counterfactuals or causality when we define  $pr\left(y | \overset{\vee}{x}, \overset{\vee}{z}, w\right)$  to be  $h\{y, w | g = (x, z)\} / h\{w | g = (x, z)\}$ . Unfortunately, since covariates are missing, an investigator must rely on often shaky subject matter beliefs to guide link-deletions. Pearl and Verma (1991) appear to argue, although I would not fully agree, that beliefs about causal associations are quite generally sharper and more accurate than those about non-causal associations. If they are correct, it would be advantageous to have all potential links on  $G$  represent direct causal effects, which will be the case only if  $V$  is a finest fully randomized causal graph and would justify Pearl's focus on non-parametric structural equation models.

### Bibliography

Dawid, A.P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series A* **41**, 1-31.

Pearl, J. & Verma, T. (1991). A theory of inferred causation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference*. Eds. J.A. Allen, R. Fikes and E. Sandewall, pp. 441-452. San Mateo, CA: Morgan Kaufmann.

Robins, J.M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**, 1393-1512.

Robins, J.M. (1987). Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect." *Computers and Mathematics with Applications*, **14**, 923-945.

Robins, J.M., Blevins, D., Ritter, G. and Wulfsohn, M. (1992). G-estimation of the effect of prophylaxis therapy for pneumocystic carinii pneumonia on the survival of AIDS patients. *Epidemiology* **3**, 319-336.

Robins, J.M. (1993). Analytic methods for estimating HIV treatment and cofactor effects. In *Methodological Issues of AIDS Mental Health Research*. Eds.

D.G. Ostrow & R. Kessler, pp. 213-290. New York, NY: Plenum Publishing.

Rosenbaum, P. & Rubin, D. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.

Rubin, D.B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* **7**, 34-58.

Spirtes, P., Glymour, C., & Schienens, R. (1993). *Causation, Prediction, and Search*. New York, NY: Springer-Verlag.