ELSEVIER

Journal of Econometrics 112 (2003) 89–106

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Commentary

General methodological considerations

James M. Robins*

Departments of Epidemiology and Biostatistics, Harvard University School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA

1. Introduction

I wish to thank Alok Bhargava for organizing a discussion of this interesting paper by Adams et al. My discussion will focus on quite generic philosophical, statistical, and modelling issues that arise when inferring causal effects from complex panel and time series data. The analysis of Adams et al. appears to be driven by the following philosophical proposition: when the sample size is large and a test of the statistical hypothesis of Granger non-causality accepts, one can validly infer lack of causation based on observational panel and/or time series data, even in the absence of subject-matter-specific background knowledge; however, because of the possibility of hidden (unmeasured) common causes, it is never possible to infer causality in the absence of background knowledge. In this discussion, I will argue that the authors' apparent philosophical belief that it is possible to infer lack of causation from lack of association in the absence of substantive subject-matter knowledge implicitly uses the "faithfulness analysis" of causation made explicit by Spirtes et al. (1993) in their book *Causation*, *Prediction and Search* and implemented in their computer program Tetrad. In Section 2, I review the assumptions underlying a "faithfulness analysis". In Section 3, I will describe some of the surprising implications of these analyses due to Spirtes et al. (1993) and Pearl and Verma (1991). Specifically, I will show that in certain specific settings and contrary to the intuitions of the authors', it is also possible to infer causality using a faithfulness analysis.

Sections 2 and 3 are devoted to describing how a faithfulness analysis can be used to infer causation and non-causation based on dependencies and independencies found in observational data in the absence of subject-matter specific background knowledge. In Section 4, I argue, following Robins (1997) and Robins and Wasserman (1999), that a faithfulness analysis is not appropriate when, as in most econometric and epidemiologic studies, the probability that any two given study variables have no common cause is close to zero. As a consequence, in most econometric and epidemiologic studies,

* Tel.: +1-617-432-0206; fax: +1-617-432-1884.
E-mail address: robins@hsph.harvard.edu (J.M. Robins).

PII: S0304-4076(02)00151-3

small causal effects can never be ruled in nor ruled out based on observational data and background substantive knowledge is required to rule in or out even fairly large causal effects. In Section 7, I recommend a sensitivity analysis approach to causal inference as an alternative to a faithfulness analysis approach. I conclude by outlining a unifying Bayesian approach that clarifies the relationship between a faithfulness analysis approach and a sensitivity analysis approach.

To motivate the remainder of my discussion, I now describe another quite surprising implication of a "faithfulness analysis" of causation that is treated in detail in Example 4 of Section 2. Suppose in the author's notation $S_t=(S_{1t},S_{2t})$ has two components, (H_t,S_{1t},S_{2t}) are temporarily ordered as indicated, and we have correctly specified parametric models for $f(H_t\,|\,\bar{S}_{t-1},\bar{H}_{t-1})$ and $f(S_{2t}\,|\,S_{1t},\bar{S}_{t-1},\bar{H}_t)$ where for any process Z_t, $\mathbf{Z}=\{Z_t;t\geqslant 0\}$ denotes the entire process, $\bar{Z}_t=(Z_0,Z_1,Z_2,\ldots,Z_t)$ denotes the history through t, and we have not imposed the Markov assumption of Adams et al. to keep the discussion more non-parametric. Suppose we fit the above models and use fitted models to carry out a life course simulation as in Adams et al. in which we intervene and set S_{1t} to s_{1t} for $t=0,1,2,\ldots$. Suppose the simulated distribution of H_t does not depend but the distribution of $S_{2(t-1)}$ does depend on the values s_{1j} to which we set S_{1j} for $j\leqslant t-1$. Robins (1986, 1989) refers to the simulated distribution of H_t as the Monte Carlo parametric g-computation algorithm estimator of the law of H_t. Then, using a faithfulness analysis of causation, we can surprisingly conclude that $\mathbf{S}=(\mathbf{S}_1,\mathbf{S}_2)$ is non-causal for $\mathbf{H}=\{H_t;t\geqslant 0\}$ even when, based on a correct model for $f(H_t\,|\,\bar{S}_{t-1},\bar{H}_{t-1})$, both (i) a test of the statistical Granger non-causality hypothesis

$$H_t \amalg \bar{S}_{t-1}\,|\,\bar{H}_{t-1} \quad \text{for all } t\geqslant 0 \tag{1}$$

rejects where $A\amalg B\,|\,C$ means A is independent of B given C and (ii) the simulated distribution of H_t depends on the value of $s_j=(s_{1j},s_{2j})$, $j<t$ to which we set $S_j=(S_{1j},S_{2j})$ in a life course simulation in which we intervened on both components of S_j. Note (i) and (ii) are logically equivalent. To summarize, a faithfulness analysis of causation implies $\mathbf{S}$ does not cause $\mathbf{H}$ if either (a) the Granger non-causality test of (1) accepts or (b) a simulated intervention on $\mathbf{S}_1=\{S_{1t};t\geqslant 0\}$ has an effect on $\mathbf{S}_2=\{S_{2t};t\geqslant 0\}$ but not on $\mathbf{H}$. We argue in Section 4 that when $\mathbf{S}$ does not cause $\mathbf{H}$, one is more likely to detect this fact by criteria (b) than by criteria (a) in the absence of measurement error in $\mathbf{S}_1$ and $\mathbf{S}_2$. In Section 5, we show this is no longer the case if $\mathbf{S}_1$ or $\mathbf{S}_2$ is measured with error.

Let 'NULL' be the set of all laws $\{f(H_{t+1}\,|\,\bar{S}_t,\bar{H}_t),\ f(S_{2t}\,|\,S_{1t},\bar{S}_{t-1},\bar{H}_t);\ t{=}0,1,2,\ldots\}$ for which (i) the Granger non-causality hypothesis (1) is false and (ii) a simulated intervention on $\mathbf{S}_1$ has an effect on $\mathbf{S}_2$ but not on $\mathbf{H}$. It is important to be able to detect whether a member of this set generated the data, because, under a faithfulness analysis of causality, this would imply no causal effect of $\mathbf{S}$ on $\mathbf{H}$. In Section 5, we show that if S_{1t} is continuous and S_{2t} is dichotomous, and we model $f(H_{t+1}\,|\,\bar{S}_t,\bar{H}_t)$ and $f(S_{2t}\,|\,S_{1t},\bar{S}_{t-1},\bar{H}_t)$ using the linear latent variable models described by Eqs. (2)–(4) of Adams et al. then, when the true data generating process lies in 'NULL,' we are guaranteed to falsely reject the hypothesis 'NULL' with probability approaching one. The reason is that the set 'NULL' and set of laws represented by Adams et al.'s linear latent variable models have an empty intersection when S_{1t} is continuous and

S_{2t} is dichotomous. It follows that when the true law is in “NULL”, the Adams et al. Models (2)–(4) must be misspecified; as a consequence an analysis based on these models will incorrectly discover an apparent simulated life course effect of $\mathbf{S}_1$ on $\mathbf{H}$ due to this model misspecification. This problem first identified in Robins (1986) has been referred to as the null paradox by Robins and Wasserman (1997).

In Section 6, we describe three alternative models for the joint distribution of $(\mathbf{H},\mathbf{S})$ that avoid the null paradox exhibited by Adams et al. Models (2)–(4): marginal structural models, structural nested models, and a modification of the linear latent variable model suggested by Cox and Wermuth (1999). Marginal structural models and structural nested models also form the basis of the sensitivity analysis approach recommended in Section 7.

2. Causation, directed acyclic graphs and faithfulness

Consider a study in which there exists a set of M temporally ordered random variables $\mathbf{V}=(V_1,\ldots,V_M)$ with density $f(\mathbf{v})$ with respect to a dominating measure μ. Here $\mathbf{v}$ is a realization of $\mathbf{V}$. Let $\mathbf{A}=\bar{A}_K=(A_0,\ldots,A_K)\subset\mathbf{V}$ be a temporally ordered subset of $\mathbf{V}$ where for any variable X_t, $\bar{X}_t=(X_0,\ldots,X_t)$ is the history of that variable up to t. Let Z_0 be all members of $\mathbf{V}$ that are temporally prior to A_0, Z_t be all members of $\mathbf{V}$ that are temporally prior to A_t but after A_{t-1}, and Z_{K+1} be all the members of V that are subsequent to A_K. Then we can write the density of $\mathbf{V}$ as $f(\mathbf{v})=\prod_{t=0}^{K+1}f(z_t\,|\,\bar{z}_t,\bar{a}_{t-1})\prod_{t=0}^{K}f(a_t\,|\,\bar{z}_t,\bar{a}_{t-1})$ with $\bar{z}_{-1}\equiv\bar{a}_{-1}\equiv 0$. Now, given a set $d=(d_0,\ldots,d_K)$ of conditional densities $d_t=d_t(a_t\,|\,\bar{z}_t,\bar{a}_{t-1})$ for $A_t\,|\,\bar{Z}_t,\bar{A}_{t-1}$, let

$$f_d(\mathbf{v})=\prod_{t=0}^{K+1}f(z_t\,|\,\bar{z}_{t-1},\bar{a}_{t-1})\prod_{t=0}^{K}d_t(a_t\,|\,\bar{z}_t,\bar{a}_{t-1}) \tag{2}$$

be the density of $\mathbf{V}$ formed by replacing the density $f(a_t\,|\,\bar{z}_t,\bar{a}_{t-1})$ that generated the data by $d_t(a_t\,|\,\bar{z}_t,\bar{a}_{t-1})$. Let $f_d^{\text{int}}(\mathbf{v})$ be the density of $\mathbf{V}$ that would result if one had intervened and, possibly contrary to fact, conducted a sequential randomized trial in which A_t had been randomly assigned with randomization probability $d_t(a_t\,|\,\bar{z}_t,\bar{a}_{t-1})$. We shall call d a treatment regime or strategy for $\mathbf{A}$ and $f_d^{\text{int}}(v)$ the causal effect of d on $\mathbf{V}$. Let $\mathscr{D}(\mathbf{A})$ be the set of regimes for $\mathbf{A}$. We write $f_d^{\text{int}}(\mathbf{x})$ and $f_d(\mathbf{x})$ for the marginal density of $\mathbf{X}\subset\mathbf{V}$ under $f_d^{\text{int}}(\mathbf{v})$ and $f_d(\mathbf{v})$. If each $d_t(a_t\,|\,\bar{z}_t,\bar{a}_{t-1})$ is a degenerate distribution putting mass 1 on $a_t=\tilde{d}_t(\bar{z}_t,\bar{a}_{t-1})$, we say the regime d is the non-random regime $\tilde{d}=(\tilde{d}_0,\ldots,\tilde{d}_K)$ characterized by the functions $\tilde{d}_0,\ldots,\tilde{d}_K$. Otherwise d is random. Let $\mathbf{a}^*=(a_0^*,\ldots,a_K^*)$ be a realization of $\mathbf{A}$. If for each t a non-random regime $\tilde{d}_t(\bar{z}_t,\bar{a}_{t-1})$ takes the value a_t^* for all $(\bar{z}_t,\bar{a}_{t-1})$, we say that the non-random regime $\tilde{d}$ is non-dynamic and write $\tilde{d}=\mathbf{a}^*$; otherwise it is dynamic.

Definition I. If, for given sets $\mathbf{A}$ and $\mathbf{X}\subset\mathbf{V}\backslash\mathbf{A}$ and regime $d\in\mathscr{D}(\mathbf{A})$, $f_d(\mathbf{x})=f_d^{\text{int}}(\mathbf{x})$, we say that, given data on $\mathbf{V}$, there is no confounding for the causal effect of $\mathbf{A}$ on $\mathbf{X}$ under regime d.

In the language of the Adams et al. paper, if the density of $\mathbf{X}$ one would obtain by simulating from $f_d(\mathbf{x})$ is the same as the intervention density $f_d^{\text{int}}(\mathbf{x})$ then there is no confounding and the intervention distribution $f_d^{\text{int}}(\mathbf{x})$ is identified from the joint distribution of the $Z_t \mid \bar{A}_{t-1}, \bar{Z}_{t-1}$, $t = 0, \ldots, K$. As Adams et al. point out, in an observational study one can never be certain that the identifiable distribution $f_d(\mathbf{x})$ equals $f_d^{\text{int}}(\mathbf{x})$ because of the possibility that there exist direct common causes of $\mathbf{A}$ and $\mathbf{X}$ that are not contained in $\mathbf{V}$. Robins (1986) referred to $f_d(\mathbf{x})$ as the g-computation algorithm formula (hereafter g-formula) for the density of $\mathbf{X}$ under regime d. Because $f_d(\mathbf{x})$ and $\mathscr{D}(\mathbf{A})$ depend on the particular variables included in $\mathbf{V}$ we write $f_d^{\mathbf{v}}(\mathbf{x})$ and $\mathscr{D}^{\mathbf{v}}(A)$ when we wish to make this dependence explicit.

Definition. If for every set $\mathbf{A} \subset \mathbf{V}$ and every regime $d \in \mathscr{D}(\mathbf{A})$, $f_d(\mathbf{v}) = f_d^{\text{int}}(\mathbf{v})$, we say that the set V is a causal system.

Definition. Given a causal system $\mathbf{V}$, we say treatment $\mathbf{A} = \bar{A}_K = (A_0, \ldots, A_K)$ is not a cause of the set of variables $\mathbf{X} \subset \mathbf{V}$ if $f_d(\mathbf{x}) = f_d^{\text{int}}(\mathbf{x})$ is the same for all $d \in \mathscr{D}(\mathbf{A})$. Otherwise, we say that $\mathbf{A}$ is a cause of $\mathbf{X}$.

Remark. A sufficient (but not necessary) condition for $\mathbf{V}$ to be a causal system is that $\mathbf{V}$ is the observed data from an underlying recursive non-parametric structural equation model (NPSEM) as in Pearl (1995) or a finest fully randomized causally interpreted structured tree graph (FRCISTG) as in Robins (1986). Indeed, a NPSEM ⇒ FRCISTG ⇒ causal system. See Robins (1995, 2002) for further discussion. Indeed, a linear recursive structural equation model (SEM) with normal errors can be interpreted as a causal system.

A causal system has an elegant and informative graphical representation using directed acyclic graphs (DAGs) that is described in Theorem 1 below. A DAG G is a set of vertices (nodes) with arrows (directed edges) between some pairs of vertices such that there are no directed cycles, i.e., one cannot start at a vertex and follow a directed path back to that vertex. A directed path from vertex A to vertex B is a sequence of vertices starting at A and ending at B of the form $A \to X_1 \to \cdots \to X_k \to B$ where the X's represent other vertices. For any node V_m, we let $PA_{m,G}$ denote the set of nodes that are parents of V_m on G (i.e., those nodes with arrows pointing directly into V_m). We associate with a DAG G and a set $\mathbf{V} = (V_1, \ldots, V_M)$ of random variables represented by the vertices of the DAG, a statistical model $\mathscr{F}(G)$ which consist of all densities $f(\mathbf{v})$ that factorize as

$$f(\mathbf{v}) = \prod_{m=1}^{M} f(v_m \mid pa_{m,G}). \tag{3}$$

A density $f(\mathbf{v})$ that is in $\mathscr{F}(G)$ is said to be represented by G. The descendants of V_m are those variables that can be reached starting from V_m by following a directed path. The ancestors of V_m are those variables which have V_m as a descendant. Pearl (1988) proves the following.

Lemma 1. *Assumption* (3) *is equivalent to the assumption that each variable V_m on the DAG G is statistically independent of its non-descendants conditional on its parents.*

Assumption (3) logically implies additional conditional and unconditional statistical independencies beyond those described in Lemma 1. The d-separation criteria of Pearl (1988) and Lauritzen et al. (1990) shows how to obtain all conditional and unconditional independencies implied by (3). Specifically (3) implies that a set of variables $\mathbf{X}$ is conditionally independent of another set of variables $\mathbf{Y}$ given a third set of variables $\mathbf{Z}$ if $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ on the graph, where d-separation is a statement about the topology of the graph. To describe d-separation we first need to define the moralized ancestral graph generated by the variables in $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$. In the following a path between two variables is any unbroken sequence of edges (regardless of the directions of the arrows) connecting the two nodes.

Definition (Lauritzen et al., 1990). The moralized ancestral graph generated by the variables in $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ is formed as follows:

(i) Remove from the DAG all nodes (and corresponding edges) except those contained in the sets $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ and their ancestors.
(ii) Next, connect by an edge every pair of nodes that share a common child.

Definition. $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ if and only if on the moralized ancestral graph generated by $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, some node in $\mathbf{Z}$ intercepts (i.e., lies on) any path between any node in $\mathbf{X}$ and any node in $\mathbf{Y}$. If $\mathbf{X}$ is not d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ we say $\mathbf{X}$ and $\mathbf{Y}$ are d-connected given $\mathbf{Z}$.

A distribution $f(\mathbf{v})$ that is represented by G is said to be faithful to G if it does not have any additional conditional and unconditional independencies beyond those implied, through the d-separation criteria, by (3). For a faithful density $f(v)$, $\mathbf{X}$ and $\mathbf{Y}$ are statistically independent given $\mathbf{Z}$ if and only if $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$. We let $\mathscr{F}_{\text{faith}}(G) \subset \mathscr{F}(G)$ be the set of faithful distributions.

Definition. We say that there is a directed path from a set of nodes $\mathbf{A}$ to set of nodes $\mathbf{X}$ on a DAG G if and only if there is a directed path from some element of $\mathbf{A}$ to some element of $\mathbf{X}$. Further, we say the ordered set $V = (V_1, \ldots, V_M)$ of vertices of a DAG G are consistent with G if no V_j is a descendant of any V_i with $i > j$.

Theorem 1 (Spirtes et al., 1993). *Suppose* $\mathbf{V}$ *is a causal system and* $f(\mathbf{v}) \in \mathscr{F}(G)$ *for some DAG G consistent with* $\mathbf{V}$. *If there is no directed path from* $\mathbf{A}$ *to* $\mathbf{X}$ *then* $\mathbf{A}$ *is not a cause of* $\mathbf{X}$. *Conversely, if* $f(\mathbf{v}) \in \mathscr{F}_{\text{faith}}(G)$, *then a directed path from* $\mathbf{A}$ *to* $\mathbf{X}$ *implies* $\mathbf{A}$ *is a cause of* $\mathbf{X}$. *That is, if* $f(\mathbf{v}) \in \mathscr{F}_{\text{faith}}(G)$, *then* $\mathbf{A}$ *is a cause of* $\mathbf{X}$ *if and only if there is a directed path from* $\mathbf{A}$ *to* $\mathbf{X}$.

Suppose $\mathbf{V}$ is a causal system and $f(\mathbf{v}) \in \mathscr{F}(G)$ for DAG G. Then we say that $\mathbf{A}$ is a direct cause of $\mathbf{X}$ (relative to the variables in G) if some member of $\mathbf{A}$ is a parent of some member of $\mathbf{X}$. We conclude from Lemma 1 that each variable in $\mathbf{V}$ is independent of all variables that it does not causally influence (i.e., its non-descendants) conditional on its direct causes (i.e., parents). Spirtes et al. (1993) refer to this statement as the causal Markov assumption.

Spirtes et al. (1993) and Pearl and Verma (1991) propose methods for drawing casual conclusions from associations found in observational data based on the following assumption.

Faithfulness assumption: The distribution of any set of observed variables $\mathbf{O}$ is the marginal distribution $f(\mathbf{o})=\int f(\mathbf{o},\mathbf{u})\,\mathrm{d}\mu(\mathbf{u})$ for $\mathbf{O}$ from a density $f(\mathbf{v})=f(\mathbf{o},\mathbf{u})$ faithful to a DAG G whose vertex set $\mathbf{V}=(\mathbf{O},\mathbf{U})$ is a causal system. This notation does not imply that $\mathbf{O}$ is temporal prior to $\mathbf{U}$.

The faithfulness assumption is referred to as the stability assumption in Pearl (2000). One justification for the faithfulness assumption is that, in cases where the model $\mathscr{F}(G)$ can be parameterized by a finite dimensional parametric family, the subset of unfaithful distributions typically has Lebesgue measure zero on the parameter space and thus is a priori unlikely. The faithfulness assumption states that we ignore the possibility that the data were generated by one of these a priori unlikely distributions. A more intuitive, philosophical justification is considered in Section 4. The faithfulness assumption is used as follows in a *faithfulness analysis of causation*.

Faithfulness analysis of causation: Suppose in an observational study the distribution $f(\mathbf{o})$ of the observed variables was known. Spirtes et al. (1993) and Pearl and Verma (1991) then show how to find all DAGs $G^{(j)}$, $j=1,2,\ldots$ with the vertex set $\mathbf{V}^{(j)}=(\mathbf{O},\mathbf{U}^{(j)})$ such that $f(\mathbf{o})$ is the marginal of a density $f(\mathbf{o},\mathbf{u}^{(j)})\in\mathscr{F}_{\text{faith}}(G^{(j)})$ and the partial ordering of $G^{(j)}$ is consistent with what is known about the temporal ordering of the elements of $\mathbf{O}$. The $\mathbf{U}^{(j)}$ are the hidden (unobserved) variables in $\mathbf{V}^{(j)}$. Next, the causal and non-causal relationships implied by the topology of $G^{(j)}$ and the assumption that $\mathbf{V}^{(j)}$ is a causal system are obtained using Theorem 1. Finally, the causal and non-causal relationships that are true in each and every $G^{(j)}$ are reported as causal and non-causal, respectively, since it is possible that any of the $\mathbf{V}^{(j)}$ could be the true causal system generating the data. No decision is offered concerning causal and non-causal relations that are true in some but not every $G^{(j)}$. The fast causal inference (FCI) algorithm of Spirtes et al. (1993) implements a faithfulness analysis of causation.

3. Examples

We now describe the results of applying a faithfulness analysis of causation to a number of examples. Examples 2–4 are the examples mentioned in the introduction. For the moment, we assume that we are given the joint distribution of the observed variables $\mathbf{O}$.

Example 1. Suppose that $O=(S,H)$ with S and H univariate and S and H are statistically independent. It follows by applying the FCI algorithm that in all the $G^{(j)}$ defined above, (i) S and H are not connected by a directed path so neither causes the other and (ii) S and H do not have a common ancestor so they do not share a common cause.

Example 2. Suppose $\mathbf{O} = (\mathbf{H}, \mathbf{S})$ where, as in the introduction $\mathbf{H} = \{H_t; t \geqslant 0\}$ and $\mathbf{S} = \{S_t; t \geqslant 0\}$ and the variables are ordered by time t and S_t follows H_t. Suppose the Granger non-causality null hypothesis (1) is true for all t. Then it follows by applying the FCI algorithm that, in all $G^{(j)}$, no element of $\mathbf{S}$ is ancestor of an element of $\mathbf{H}$ (so $\mathbf{S}$ does not cause $\mathbf{H}$) and whenever $t < t'$, S_t and $H_{t'}$ have no unmeasured common direct cause (i.e., all common parents are in $\mathbf{O}$).

Example 3. Suppose (S_1, S_2, H) are univariate and temporally ordered as indicated, and $S_1 \amalg H \mid S_2$ is the only conditional or unconditional independence satisfied by their law. Then it follows by applying the FCI algorithm that, in all $G^{(j)}$, S_2 is a cause of H, S_1 is not a cause of H, S_2 and H have no common cause, and S_1 and H have no common cause. Hence our faithfulness analysis has succeeded in inferring S_2 is a cause of H in the absence of any substantive background knowledge.

Example 4. Suppose $\mathbf{O} = (\mathbf{H}, \mathbf{S})$ where, as in the introduction $\mathbf{H} = \{H_t; t \geqslant 0\}$ and $\mathbf{S} = \{S_t; t \geqslant 0\} = (\mathbf{S}_1, \mathbf{S}_2)$ and the variables are ordered by time t and H_t precedes S_{1t} which precedes S_{2t}. We need the following definition concerning sets of variable $\mathbf{A}, \mathbf{X}, \mathbf{O}$ with $\mathbf{A} \cap \mathbf{X} = \varnothing$ and $\mathbf{A} \cup \mathbf{X} \subset \mathbf{O}$.

Definition (Robins, 1986). The "g"-null hypothesis for the effect of $\mathbf{A}$ on $\mathbf{X}$ based on $\mathbf{O}$ holds if the g-formula density $f_d^{\mathbf{o}}(\mathbf{x})$ for $\mathbf{X}$ based on the variables in $\mathbf{O}$ is the same for all regimes $d \in \mathscr{D}^{\mathbf{o}}(A)$.

Suppose that the Granger null hypothesis (1) is false. Further suppose that the "g"-null hypothesis based on $\mathbf{O}$ holds for the effect of $\mathbf{S}_1$ on $\mathbf{H}$ but not for the effect of $\mathbf{S}_1$ on $\mathbf{S}_2$. This latter supposition is equivalent to the supposition that a simulated intervention on $\mathbf{S}_1$ has an effect on $\mathbf{S}_2$ but not on $\mathbf{H}$. The "g-null" theorem of Robins (1986) states that the "g"-null hypothesis for the effect of $\mathbf{S}_1$ on $\mathbf{H}$ based on $\mathbf{O}$ holds if and only if

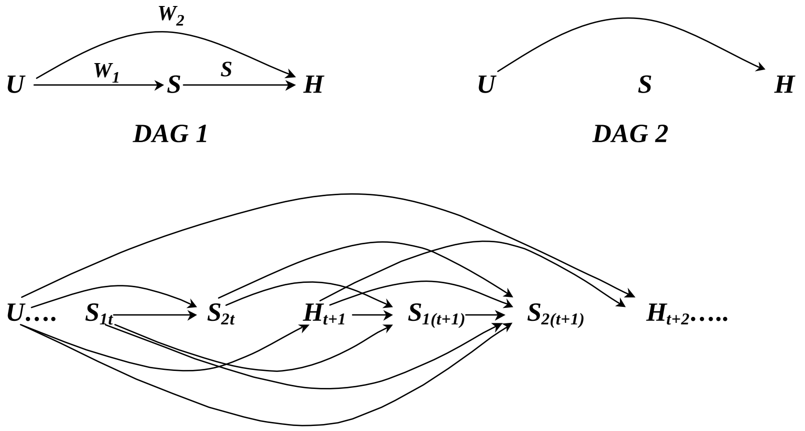
$$S_{1t} \amalg (H_{t+1}, H_{t+2}, \ldots) \mid \bar{S}_{t-1}, \bar{H}_t, t \geqslant 0. \tag{4}$$

We assume that (4) represents the only conditional or unconditional independences satisfied by $f(\mathbf{O})$. Then it follows by applying the FCI algorithm that, in all $G^{(j)}$, no element of $\mathbf{S}$ is ancestor of an element of $\mathbf{H}$ (so $\mathbf{S}$ does not cause $\mathbf{H}$) and whenever $t < t'$, S_{1t} and $H_{t'}$ have no unmeasured common direct cause (i.e., all common parents are in $\mathbf{O}$).

In practice, the law of $\mathbf{O}$ is unknown; instead we might observe $(\mathbf{O}_1, \ldots, \mathbf{O}_n), n$ i.i.d. draws from the unknown law of $\mathbf{O}$. In that case, Spirtes et al. (1993) suggest that when a p-value for a test of a particular conditional or unconditional independence is sufficiently large (say > 0.5) and the sample size n is large, one can conclude independence holds in the population and apply the above faithfulness analysis.

4. Intuitive justification for a faithfulness analysis and a critique

In this section, I give the intuitive justification offered by Pearl and Verma (1991) and Spirtes et al. (1993) for a faithfulness analysis in the context of the above



DAG 3: U is prior to H_o
DAG 3 is complete except (1) S_{1j} is not a child of U for all j.
(2) H_j is not a child of any S_{1t} or S_{2t} for all j,t.

Fig. 1. Some casual graphs.

examples. I also describe the objections to this justification raised by myself and Larry Wasserman.

Consider again Example 1 of Section 3. Let DAG 1 in Fig. 1 represent a causal system $V = (U,S,H)$ where U includes all unmeasured potential common causes of a particular social factor S and heath outcome H. To ease the exposition, we assume for the moment U is unidimensional and DAG 1 represents a linear recursive SEM in which all disturbances (error terms) are normal with mean zero, and U, S and H have variance one. The lower case letters in DAG 1 denote path coefficients. If, as a fact of nature, a path coefficient is zero, then the corresponding arrow is to be regarded as "missing". For example, if S causes H, the path coefficient s differs from zero. In contrast, if S does not cause H, s equals zero and the arrow from S to H can be removed.

We first assume that we know the law of $O = (S,H)$ and that $S \coprod H$. Spirtes et al. (1993) then conclude that S does not cause H (i.e., $s = 0$) as follows. The only way we could have $S \coprod H$ is for one of the following to hold.

Explanation (1). U is a common cause ($u_1 u_2 \neq 0$) and S causes $H(s \neq 0)$, but the magnitudes of the path coefficients (u_1, u_2, s) are perfectly balanced (i.e., $s - u_1 u_2 = 0$) in such a way to make $S \coprod H$.

Explanation (2). U is not a common cause ($u_1 u_2 = 0$) and S does not cause $H(s = 0)$ so the arrow from S to H and at least one of the arrows from U are erased resulting in DAG 2, say, shown in Fig. 1.

Spirtes et al. (1993) then proceed to rule out Explanation (1) (which is always a logical mathematical possibility) by the following "faithfulness" argument: Since the subset $\{(u_1, u_2, s);\ u_1 u_2 - s = 0\}$ of path coefficients that will make $S \coprod H$ when

$u_1u_2 \neq 0$ and $s \neq 0$ has Lebesgue measure zero in R^3, the "prior" probability of Explanation (1) occurring is zero (since Explanation (1) requires such "fortuitous" values of the path coefficients). Thus, we are left with Explanation (2) and the corresponding graph DAG 2. Note since $S \coprod H$, we refer to this argument as a faithfulness argument because the rejected Explanation (1) implies that the law of $V = (U, S, H)$ is not faithful to any DAG 1, while the accepted Explanation (2) implies that the law of V is faithful to DAG 2. In summary, we have gone from no population association to inferring no causation in the absence of any subject matter specific knowledge by using a faithfulness argument; indeed, we did not need knowledge of the real world variables represented by S and H.

I fully agree with Spirtes et al. (1993) that the prior probability of Explanation (1) is zero. However, Spirtes et al. (1993) then conclude that Explanation (2) must be the proper explanation and thus conclude S does not cause H. It is this last step with which I disagree. This step can only be justified if the prior probability that U is not a common cause ($u_1u_2 = 0$) is greater than zero. Otherwise, both Explanations (1) and (2) are events of probability zero and so their relative likelihood is not defined without further assumptions. Now the Lebesgue measure in R^2 of the event "$u_1u_2 = 0$" is also zero. It follows that Spirtes et al. (1993) are making the "hidden" assumption that the event that a path coefficient has value zero (i.e., that an arrow is missing on the graph) has a non-zero probability. On this "hidden" assumption, Explanation (2) does have a positive probability, and we can conclude that S does not cause H. Now I do agree that, for any two given variables, the probability that the path coefficient between them is zero (i.e., that neither variable causes the other) is non-zero. But in the epidemiologic example in which we are trying to determine whether a particular S causes a particular heath outcome H, if, as assumed, $V = (U, S, H)$ is a causal system, then the unmeasured variable U on DAG 1 does not represent the single variable we have been assuming for expositional convenience, but rather represents all possible unmeasured common causes of S and H. It follows that if there exists any common cause of S and H, Explanation (2) is false.

Now I (and every other epidemiologist I know) believe that, given a sufficiently large sample size, a test for independence between any two variables recorded in an observational study will reject due to the existence of unmeasured common causes (confounders). Thus, we teach epidemiology students that, with certainty (i.e., with prior probability 1), Explanation (2) will not occur (once we understand that U represents the conglomeration of all unmeasured potential common causes). It follows that we cannot conclude Explanation (2) is to be preferred to Explanation (1). Thus, we cannot conclude that S does not cause H.

Suppose, however, that, although Explanations (1) and (2) both have zero prior probability and one believes under further reasonable assumptions and with respect to some measure, the relative likelihood of Explanation (1) is infinitely greater than that of Explanation (2). We now argue that even so, nonetheless, in realistic studies with their finite sample size, we should still not invoke "a faithfulness analysis" and conclude that S does not cause H when S and H are uncorrelated in the data.

Consider a realistic study, even a very large one, with its finite sample size. Then, even if, in the data, the sample correlation $\hat{\rho}$ between S and H is exactly zero and the p-value is precisely 1.0, nonetheless, due to sampling variability, there will exist a confidence interval around the null estimate of the association parameter ρ between S and H. Since, as argued above, our prior probability that in truth S and H are independent is zero, (as Explanations (1) and (2) both have prior probability zero), our *posterior* odds of the hypothesis that the population correlation ρ is exactly zero will be zero compared to the hypothesis that S and H are dependent with their ρ lying in the small confidence interval surrounding zero. Thus, again, we would not necessarily conclude that S does not cause H since the event that we "almost" have balanced confounding ($s - u_1u_2 \neq 0$ but small) need not have a prior probability of near 0 and thus its posterior probability need not be small. However, if, based on subject matter knowledge, we believed that the magnitude of the confounding by the conglomerate variable U was small ($u_1u_2 \neq 0$ but small) and that s had a reasonable prior probability of being 0, then we would conclude, based on a formal Bayesian analysis, that the posterior probability that s was 0 exceeded $\frac{1}{2}$ and thus that S likely did not cause H. This analysis would, of course, reflect our subject matter knowledge and would not be based on the subject-matter-independent philosophical principle of a faithfulness analysis. Humphreys and Freedman (1996) make a closely related argument.

More generally, Robins and Wasserman (1999) prove that under an asymptotics for which the probability of no unmeasured common causes is small relative to sample size neither causal nor non-causal relationships are identifiable from observational data in the absence of substantive background knowledge, even if the faithfulness assumption is true. They also argue that such an asymptotics accurately captures the beliefs of most practicing epidemiologists.

In most observational studies there will exist measured pretreatment potential common causes. The above critique applies to this case by arguing conditionally on these measured variables. Indeed, as emphasized by Greenland and Robins (1986), Pearl and Verma (1991), and Spirtes et al. (1993), conditioning on pretreatment variables has the potential to increase rather than decrease the degree of selection bias. Further, in my opinion, there will always remain unmeasured common causes.

We next consider the implications of the above justification for and critique of a faithfulness analysis for the other examples of Section 3. Consider first Example 3. In Robins (1997, Section 11), I describe Pearl and Verma's (1991) and Spirtes et al.'s (1993) intuitive justification for and my critique of the conclusion that one can infer that S_2 causes H in Example 3 from associations found in the data. My exposition there so closely tracts the exposition just given that I am near certain that any reader will acknowledge that (i) if one agrees with Spirtes et al. (1993) that one can infer non-causality in Example 1 in the absence of substantive knowledge then one can infer causality in Example 3 and (ii) if one agrees with me that one cannot infer non-causality in Example 1 in the absence of substantive knowledge then one cannot infer causality in Example 3.

Turn now to Example 2. The only difference between Example 2 and that discussed in the last paragraph of my critique of Example 1 is quantitative in that we have repeated tests (one for each t) of the Granger null hypothesis (1). Based on my critique under Example 1, I do not believe Adams et al. are justified in inferring non-causality when the tests of (1) do not reject without first arguing quantitatively, based on subject matter knowledge, about the likelihood that a non-zero causal effect of $\mathbf{S}$ on $\mathbf{H}$ is nearly balanced by an association induced by unmeasured common causes. Furthermore, if the authors do disagree with my critique and believe they are still justified in inferring non-causality in Example 2, then I believe that, to be logically consistent, they must also believe that one can infer causality in the setting of Example 3 and non-causality in the setting of Example 4, as both conclusions are based on the same faithfulness analysis that is required to justify the conclusion of Example 2.

Next consider Example 4. The distribution $f(\mathbf{O})$ of Example 4 is the marginal of a distribution that is faithful to DAG 3 of Fig. 1. Faithfulness follows from the fact that the only conditional independences among the variables in $\mathbf{O}$ implied by G is Eq. (4). Let DAG Granger be DAG 3 with all arrows from U to S_2 removed. The distribution $f(\mathbf{O})$ of Example 2 is the marginal of a distribution that is faithful to DAG Granger since the only conditional independences among the variables in $\mathbf{O}$ implied by DAG Granger is (1). If the vertex set $\mathbf{V} = (U, \mathbf{O})$ is a causal system, then for both DAG 3 and DAG Granger, we see from Theorem 1 that $\mathbf{S}{=}(\mathbf{S_1}, \mathbf{S_2})$ does not cause $\mathbf{H}$. Because DAGs 3 and Granger are representative of all DAGS $G^{(j)}$ obtained in a faithfulness analysis of Examples 4 and 2, respectively, $\mathbf{S} = (\mathbf{S_1}, \mathbf{S_2})$ does not cause $\mathbf{H}$ in either example under a faithfulness analysis. Intuitively the reason we could infer no effect of $\mathbf{S} = (\mathbf{S_1}, \mathbf{S_2})$ on $\mathbf{H}$ in DAG 3 is that S_{1t} is acting like an instrumental variable for S_{2t}: S_{1t} itself is correlated with S_{2t}, does not directly cause $\mathbf{H}$, and is unconfounded in the sense that there is no unmeasured common direct cause of S_{1t} and $\mathbf{H}$. In DAG Granger, both S_{1t} and S_{2t} are unconfounded in the sense that there is no unmeasured common direct cause of either S_{1t} or S_{2t} and $\mathbf{H}$.

Suppose one disagrees with my view and believes the prior probability of no unmeasured common direct causes, though quite small, is not too small to invalidate a faithfulness analysis of causation. Even so, one would expect that it is much more likely that only S_{1t} and $\mathbf{H}$ have no unmeasured common direct cause than that both S_{1t} and S_{2t} have no unmeasured common direct cause with $\mathbf{H}$. Thus tests of the Granger null hypothesis (1) should succeed much less often than tests of the null hypothesis (4) (or equivalently tests of the "g"-null hypothesis that simulated intervention on $\mathbf{S}_1$ has an effect on $\mathbf{S}_2$ but not on $\mathbf{H}$) in 'discovering' $\mathbf{S} = (\mathbf{S_1}, \mathbf{S_2})$ does not cause $\mathbf{H}$. Actually, as shown in the next section, this argument only holds when, as we have assumed thus far, neither $\mathbf{S_1}$ nor $\mathbf{S_2}$ is measured with error.

In summary, I believe the conclusions drawn in Examples 1–4 using a faithfulness analysis all stand or fall together. Further, it is my opinion they should fall. An alternative approach to causal inference that I endorse is described in Section 7.

5. Measurement error

Adams et al. recognize that measurement error is a threat to the validity of causal analyses. We now discuss the effect of random measurement error on a faithfulness analysis of causality in Examples 1–4. We shall suppose all variables are continuous and for any univariate variable Z_t, let $Z_t^* = Z_t + \varepsilon$ where ε denotes mean zero measurement error that is independent of Z_t, all other variables Y_t, and across subject and time. In Example 1, if we observe (S^*, H^*) rather than (S, H), we will still correctly infer that S does not cause H since $S \amalg H$ implies $S^* \amalg H^*$. In Example 2, if $\mathbf{S}$ is measured with error but $\mathbf{H}$ is not so we observe $(\mathbf{S}^*, \mathbf{H})$, we will still correctly infer $\mathbf{S}$ does not cause $\mathbf{H}$ since (1) implies $H_t \amalg \bar{S}^*_{t-1} \mid \bar{H}_{t-1}$. But if $\mathbf{H}$ is measured with error we will fail to conclude that $\mathbf{S}$ does not cause $\mathbf{H}$ since (1) does not imply $H^*_t \amalg \bar{S}_{t-1} \mid \bar{H}^*_{t-1}$. In Example 4, we will fail to correctly conclude that $\mathbf{S}$ does not cause $\mathbf{H}$ if any of $\mathbf{S}_1, \mathbf{S}_2$ or $\mathbf{H}$ is measured with error since (4) does not imply (i) $S^*_{1t} \amalg (H_{t+1}, H_{t+2}, \ldots) \mid \bar{S}^*_{1(t-1)}, \bar{S}_{2(t-1)}, \bar{H}_t$, (ii) $S_{1t} \amalg (H_{t+1}, H_{t+2}, \ldots) \mid \bar{S}_{1(t-1)}, \bar{S}^*_{2(t-1)}, \bar{H}_t$, or (iii) $S_{1t} \amalg (H^*_{t+1}, H^*_{t+2}, \ldots) \mid \bar{S}_{1(t-1)}, \bar{S}_{2(t-1)}, \bar{H}^*_t$. As discussed in Robins (1987), condition (i) implies that, even in the absence of model misspecification, a simulated intervention on $\mathbf{S}^*_1$ based on data $\mathbf{O}^* = (\mathbf{S}^*_1, \mathbf{S}_2, \mathbf{H})$ will show an "effect" on $\mathbf{H}$ (i.e., the g-formula density $f^{\mathbf{o}^*}_d(\mathbf{h})$ will depend on $d \in D^{\mathbf{o}^*}(\mathbf{S}^*_1)$). In other words, random measurement error in the treatment $\mathbf{S}_1$ will lead to one to incorrectly conclude that $\mathbf{S}_1$ has an effect on $\mathbf{H}$ from a simulated intervention on $\mathbf{S}_1$. This example provides an interesting exception to the 'rule' that random mismeasurement of treatment does not lead to bias under the null hypothesis of no treatment effect. The reason for the exception is that S_{1t} not only functions as a treatment at time t but also as a measured proxy for an unmeasured causal confounder U_{1t} for the effect of the treatment at time $S_{1(t+1)}$ given at $t+1$. However, when mismeasured, S_{1t} is unavailable to control confounding by the unmeasured causal factor U_{1t}, resulting in bias even under the null hypothesis of no effect of $S_{1(t+1)}$.

Finally, in Example 3 we will correctly conclude that S_2 causes H when S_1 and H are mismeasured but not when S_2 is mismeasured since $S_1 \amalg H | S_2$ implies $S^*_1 \amalg H^* | S_2$ but not $S_1 \amalg H | S^*_2$.

6. The null paradox

Consider a simplified version of Example 4, where we observe $\mathbf{O} = (S_{1t}, S_{2t}, S_{1(t+1)}, H_{t+1})$ in this temporal order at time $t = 0$ only. That is, we are assuming H_t and $S_{2(t+1)}$ are each zero with probability 1, as this results in less cluttered notation without affecting the following argument. We assume that S_{1t}, $S_{1(t+1)}$, and H_{t+1} are continuous variables but S_{2t} is dichotomous. Specializing the assumptions of Example 4 to this setting, we shall assume that the Granger non-causality hypothesis (1) is false in the sense that H_{t+1} is not independent of $(S_{1t}, S_{2t}, S_{1(t+1)})$. Further, we assume that a simulated intervention on S_1 affects S_2 but not H, which implies that $f^{\mathbf{o}}_{d=(s_{1t}, s_{1(t+1)})}(s_{2t} = 1) \equiv pr[S_{2t} = 1 \mid S_{1t} = s_{1t}]$ depends on the value s_{1t} that we set in the intervention but that $f^{\mathbf{o}}_{d=(s_{1t}, s_{1(t+1)})}(h_{t+1}) \equiv \sum_{s_{2t}=0}^{1} f(h_{t+1} \mid s_{1t}, s_{2t}, s_{1(t+1)}) f(s_{2t} \mid s_{1t})$ does not depend on $(s_{1t}, s_{1(t+1)})$. Now suppose we use the simplest version of Adams et al.'s Models (2)–(4). That is, we model $H_{t+1} \mid S_{1t}, S_{2t}, S_{1(t+1)}$ as $N(\beta_0 + \beta_1 S_{1t} + \beta_2 S_{2t} + \beta_3 S_{1(t+1)}, \sigma^2)$ and we model S_{2t} as $I(S^*_{2t} > 0)$ where $S^*_{2t} \mid S_{1t}$ is $N(\alpha_0 + \alpha_1 S_{1t}, 1)$ so that $pr[S_{2t} = 1 \mid S_{1t}] = \Phi[\alpha_0 + \alpha_1 S_{1t}]$ with Φ the $N(0,1)$ cumulative distribution function.

We now demonstrate that this model is incompatible with the assumptions of Example 4. First, by S_{1t} having an intervention effect on S_{2t}, we conclude $\alpha_1 \neq 0$. Next, we compute the mean of H_{t+1} under the distribution $f^{\mathbf{o}}_{d=(s_{1t},s_{1(t-1)})}(h_{t+1})$ to obtain

$$E^{\mathbf{o}}_{d=(s_{1t},s_{1(t+1)})}(H_{t+1}) \equiv \sum_{s_{2t}=0}^{1} E[H_{t+1} \mid s_{1t}, s_{2t}, s_{1(t+1)}] f(s_{2t} \mid s_{1t})$$

$$= \beta_0 + \beta_1 s_{1t} + \beta_2 \Phi(\alpha_0 + \alpha_1 s_{1t}) + \beta_3 s_{1(t+1)}, \tag{5}$$

which, by assumption, does not depend on s_{1t} or $s_{1(t+1)}$. Therefore, since α_1 is non-zero and $\Phi(\cdot)$ is non-linear, we must have $\beta_1 = \beta_2 = \beta_3 = 0$. But all three betas cannot be zero, since then the Granger non-causality hypothesis $H_{t+1} \amalg (S_{1t}, S_{2t}, S_{1(t+1)})$ would hold. We conclude that if we use Adams et al.'s Models (2)–(4), then, in large samples we are certain to reject the true "*g*"-null hypothesis that a simulated intervention on $\mathbf{S}_1$ does not affect $\mathbf{H}$ due to model misspecification. We now discuss alternative ways to model the data $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H})$ so the "*g*"-null hypothesis is not excluded a priori.

6.1. Models that avoid the null paradox

Our first approach to avoiding the null paradox is based on a small but important modification of Adam's et al.'s Models (2)–(4) proposed by Cox and Wermuth (1999). Specifically, we again model S_{2t} as $I(S^*_{2t} > 0)$ where $S^*_{2t} \mid S_{1t}$ is $N(\alpha_0 + \alpha_1 S_{1t}, 1)$, but we now model $H_{t+1} \mid S_{1t}, S_{1(t+1)}, S^*_{2t}$ as $N(\beta_0 + \beta_1 S_{1t} + \beta_2 S^*_{2t} + \beta_3, \sigma^2)$ so that now, for example, $E[H_{t+1} \mid S_{1t}, S_{1(t+1)}, S_{2t}] = \beta_0 + \beta_1 S_{1t} + \beta_2 E[S^*_{2t} \mid S_{1t}, S_{1(t+1)}, S_{2t}] + \beta_3 S_{1(t+1)}$. To evaluate $f^{\mathbf{o}}_{d=(s_{1t},s_{1(t+1)})}(h_{t+1})$ we now also require a model for $S_{1(t+1)} \mid S_{1t}, S^*_{2t}, S_{2t}$ which we take as $N(\eta_0 + \eta_1 S_{1t} + \eta_2 S_{2t}, 1)$. We can avoid some lengthy algebra by noting that Theorem F1 of Robins (1986) or Theorem 1 of Pearl (1995) and Robins (1995) imply that if $S_{1(t+1)} \mid S_{1t}, S^*_{2t}, S_{2t}$ only depends on S^*_{2t} through S_{2t} then $f^{\mathbf{o}}_{d=(s_{1t},s_{1(t+1)})}(h_{t+1})$ equals

$$f^{\mathbf{o}^*}_{d=(s_{1t},s_{1(t+1)})}(h_{t+1}) \equiv \int f[h_{t+1} \mid s_{1t}, s^*_{2t}, s_{1(t+1)}] f(s^*_{2t} \mid s_{1t}) ds^*_{2t}, \tag{6}$$

where $\mathbf{O}^* = (S_{1t}, S^*_{2t}, S_{1(t+1)}, H_{t+1})$. Now because all the regressions in (6) are linear and the errors are normal, the null paradox is avoided. Specifically, $f^{\mathbf{o}^*}_{d=(s_{1t},s_{1(t+1)})}(h_{t+1})$ is normal with variance not depending on s_{1t} or $s_{1(t+1)}$ and mean $\beta_0 + \beta_1 s_{1t} + \beta_2(\alpha_0 + \alpha_1 s_{1t}) + \beta_3 s_{1(t+1)} = (\beta_0 + \beta_2 \alpha_0) + (\beta_1 + \beta_2 \alpha_1) s_{1t} + \beta_3 s_{1(t+1)}$. Thus, $f^{\mathbf{o}^*}_{d=(s_{1t},s_{1(t+1)})}(h_{t+1})$ does not on s_{1t} or $s_{1(t+1)}$ if and only if $\beta_3 = \beta_1 + \beta_2 \alpha_1 = 0$ which is allowed as it does not imply the Granger null hypothesis. Cox and Wermuth's approach straightforwardly extends to the full version of Example 4. Of course, Cox and Wermuth's approach is quite restrictive because of the assumption of linearity and normality. We therefore now consider two less parametric approaches.

Robins (1989, 1994, 1999) and Robins et al. (1992) have developed two classes of causal inference models for panel and time series data that do not suffer from the null paradox, the structural nested models (SNMs) and the marginal structural models (MSMs). In this discussion, we treat structural nested mean models (SNMMs) and marginal structural mean models (MSMMs) for estimating the effect of $\mathbf{S}_1$ on the mean of $\mathbf{H}$ from n i.i.d. copies $\mathbf{O}_i$ of $\mathbf{O} = \{O_t; t \geqslant 0\}, O_t = (H_t, S_{1t}, S_{2t})$, and again H_t is prior to S_t. MSMMs are generally used to estimate the effect of non-dynamic interventions

$\tilde{d}=\mathbf{S}_1$ while SNMMs are used to estimate the effect of non-random dynamic regimes $\tilde{d}$, where we have used the notation introduced in Section 2 with $\mathbf{S}_1$ in the role of $\mathbf{A}$. We shall also need to define the related pseudo-marginal structural models and pseudo-structural nested models for reasons that will become clear presently. A MSMM specifies a model

$$E^{\text{int}}_{d=\mathbf{s}_1}[H_t]=g_t(\bar{s}_{1,t-1},\psi^*),\quad \psi^*=(\psi^*_0,\psi^*_1) \tag{7}$$

for the marginal mean of H_t under the intervention density $f^{\text{int}}_{d=\mathbf{s}_1}[\mathbf{o}]$, where ψ^* is an unknown parameter to be estimated and $g_t(\bar{s}_{1,t-1},(\psi_0,\psi_1))$ is a known function satisfying $g_t(\bar{s}_{1,t-1},(\psi_0,0))$ is a constant not depending on $\bar{s}_{1,t-1}$, so $\psi^*_1=0$ represents the null hypothesis of no effect of the non-dynamic regimes $\mathbf{s}_1$ on the mean of $\mathbf{H}$. A pseudo-MSMM model replaces the non-identifiable intervention density $f^{\text{int}}_{d=\mathbf{s}_1}[\mathbf{o}]$ with the identifiable g-functional density $f^{\mathbf{o}}_{d=\mathbf{s}_1}[\mathbf{o}]$ based on the observed data $\mathbf{O}$ to obtain the pseudo-MSMM model

$$E^{\mathbf{o}}_{d=\mathbf{s}_1}[H_t]=g_t(\bar{s}_{1,t-1},\psi^*). \tag{8}$$

Models (7) and (8) will be identical if $f^{\text{int}}_{d=\mathbf{s}_1}[\mathbf{o}]=f^{\mathbf{o}}_{d=\mathbf{s}_1}[\mathbf{o}]$, as would be the case if, for example, a causal system $\mathbf{V}=(U,\mathbf{O})$ represented by DAG 3 generated the data. Until Section 7, we restrict attention to the identifiable pseudo-MSMM model (8) as it, rather than MSMM (7), can be used in a faithfulness analysis of Example 4 that does not suffer from the null paradox. Suppose we choose $g_t(\bar{s}_{1,t-1},\psi)=(1,\bar{S}'_{1,t-1})\psi=\psi_0+\sum_{m=0}^{t-1}\psi_{1m}s_{1m}$. Then a test of the hypothesis $\psi^*_1=0$ provides a test of the "g"-null hypothesis for the effect of $\mathbf{S}_1$ on $\mathbf{H}$ based on $\mathbf{O}$ (equivalently a test of Eq. (4)), as the "g"-null hypothesis implies $\psi^*_1=0$. We can estimate ψ^* by the semiparametric inverse probability of treatment weighted (IPTW) least-squares estimator $\hat{\psi}_{\text{IPTW}}$ that solves the weighted least-squares normal equation $0=\sum_{i=1}^{n}U_{\text{MSM},i}(\psi)$. Here $U_{\text{MSM}}(\psi)=\sum_t W_t[H_t-(1,\bar{S}'_{1,t-1})\psi](1,\bar{S}'_{1,t-1})'=0$

where

$$W_t=\frac{\prod_{j=0}^{t-1}f[S_{1j}\mid\bar{S}_{1,j-1}]}{\prod_{j=0}^{t-1}f[S_{1j}\mid\bar{S}_{1,j-1},\bar{H}_{1,j},\bar{S}_{2,j-1}]}.$$

W_t is informally the product over time of the ratio of the probability of receiving the treatment S_{1j} one did indeed receive given past treatment history $\bar{S}_{1,j-1}$ divided by the probability of receiving the treatment S_{1j} one did indeed receive given past treatment history $\bar{S}_{1,j-1}$ and outcome history $(\bar{H}_{1,j},\bar{S}_{2,j-1})$. Parametric models fit by maximum likelihood are used to estimate the numerator and denominator of W_t. Consistency of $\hat{\psi}_{\text{IPTW}}$ for the pseudo-SNMM parameter depends on having a correctly specified model for $f[S_{1j}\mid\bar{S}_{1,j-1},\bar{H}_{1,j},\bar{S}_{2,j-1}]$ but not for $f[S_{1j}\mid\bar{S}_{1,j-1}]$.

It will be convenient to define a pseudo-SNMM now and withhold defining a SNMM until Section 7. An additive pseudo-SNMM is a parametric model for the contrast

$$\gamma_{tk}(\bar{s}_{1t},\bar{h}_t,\bar{s}_{2(t-1)})\equiv E^{\mathbf{o}}_{d=(\bar{s}_{1t},0)}[H_k\mid\bar{h}_t,\bar{s}_{2(t-1)}]-E^{\mathbf{o}}_{d=(\bar{s}_{1(t-1)},0)}[H_k\mid\bar{h}_t,\bar{s}_{2(t-1)}],\quad k>t,$$

where $d=(\bar{s}_{1t},0)$ is the non-dynamic regime in which we set $\mathbf{S}_1$ to $\bar{s}_{1t}$ through t and to zero thereafter, "zero" is a baseline value of treatment, and the conditional expectations are under law $f^{\mathbf{o}}_{d=(\bar{s}_{1t},0)}[\mathbf{o}]$ and $f^{\mathbf{o}}_{d=(\bar{s}_{1(t-1)},0)}[\mathbf{o}]$, respectively. Note that we do not have to add the $\bar{S}_{1t}$ history to the conditioning events, since the history is non-random under the intervention history determined by these d's. When $f_d^{\mathbf{o}}[\mathbf{o}] = f_d^{\text{int}}[\mathbf{o}]$ this contrast represents the effect of one last blip of treatment of magnitude s_{1t} at time t compared to zero among people with previous treatment $\bar{s}_{1(t-1)}$ and response $(\bar{h}_t, \bar{s}_{2(t-1)})$. Now, $\gamma_{tk}(\bar{s}_{1t}, \bar{h}_t, \bar{s}_{2(t-1)}) = 0$ if $s_{1t} = 0$ or if the "g"-null hypothesis for the effect of $\mathbf{S}_1$ on $\mathbf{H}$ based on $\mathbf{O}$ holds. A pseudo-SNMM specifies that $\gamma_{tk}(\bar{s}_{1t}, \bar{h}_t, \bar{s}_{2(t-1)}) = \gamma_{tk}(\bar{s}_{1t}, \bar{h}_t, \bar{s}_{2(t-1)}, \psi^*)$ where ψ^* is an unknown parameter to be estimated and $\gamma_{tk}(\bar{s}_{1t}, \bar{h}_t, \bar{s}_{2(t-1)}, \psi)$ is a known function satisfying $\gamma_{tk}(\bar{s}_{1t}, \bar{h}_t, \bar{s}_{2(t-1)}, 0) = 0$ so that a test of the hypothesis $\psi^* = 0$ is a test of the "g"-null hypothesis. A consistent asymptotically normal semiparametric estimator of ψ^* is given by a so-called g-estimator $\hat{\psi}$ that solves $0 = \sum_i U_{\text{SNMM},i}(\psi)$ where $U_{\text{SNMM}}(\psi) = \sum_t \sum_{k:k>t} [H_k - \sum_{m=t}^{k-1} \gamma_{mk}(\bar{S}_{1m}, \bar{H}_m, \bar{S}_{2(m-1)}, \psi)]$ $\{m_t[\bar{S}_{1t}, \bar{H}_t, \bar{S}_{2(t-1)}] - E[m_t[\bar{S}_{1t}, \bar{H}_t, \bar{S}_{2(t-1)}] \mid \bar{S}_{1(t-1)}, \bar{H}_t, \bar{S}_{2(t-1)}]\}$ with m_t being a user-supplied vector function of the dimension of ψ and the expectation in $U_{\text{SNMM}}(\psi)$ is estimated by fitting a parametric model for $S_{1t} | \bar{S}_{1(t-1)}, \bar{H}_t, \bar{S}_{2(t-1)}$.

Robins (1994) and Robins et al. (1999) derive variation independent parameterizations for the joint distribution of the observed data $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{H})$ compatible with a given pseudo-MSMM or given pseudo-SNMM that avoid the null paradox. These parameterizations allows Bayesian or parametric likelihood based estimation of ψ^*, $E_d^{\mathbf{o}}[\mathbf{H}]$ and $f_d^{\mathbf{o}}[\mathbf{h}]$ for any $d \in \mathscr{D}^{\mathbf{o}}(\mathbf{S}_1)$ when desired.

7. Sensitivity analysis: an alternative to a faithfulness analysis

In this section, I describe sensitivity analyses based on MSMMs or SNMMs as an alternative to a faithfulness analysis of causation.

Sensitivity analysis for MSMMs: To describe the proposed methodology, I need to define counterfactual responses. Let $H_{d,t}$ be a subject's response H_t if, possibly contrary to fact, the subject followed the regime $d \in \mathscr{D}^{\mathbf{o}}(\mathbf{S}_1)$. By assumption $H_{d=\mathbf{s}_1,t} = H_{d=(\bar{s}_1(t-1).0),t}$ and $H_t = H_{d=(\bar{S}_1(t-1).0),t}$. When $f_d^{\mathbf{o}}(\mathbf{o}) = f_d^{\text{int}}(\mathbf{o})$ for all $d \in \mathscr{D}^{\mathbf{o}}(\mathbf{S}_1)$, then, (i) by definition there are no unmeasured confounders for the effect of $\mathbf{S}_1$ on $\mathbf{H}$ given data on $\mathbf{O}$, and (ii) the contrast $E[H_{d=\mathbf{s}_1,t} \mid \bar{h}_m, \bar{s}_{m-1}, s_{1m}] - E[H_{d=\mathbf{s}_1,t} \mid \bar{h}_m, \bar{s}_{m-1}, s_{1m}^{\dagger}]$ between the conditional counterfactual mean among subjects with $S_{1m} = s_{1m}$ with that among subjects with $S_{1m} = s_{1m}^{\dagger}$ with a common observed past $(\bar{h}_m, \bar{s}_{m-1})$ takes the value of zero. Hence, a natural approach to modelling the magnitude of confounding of the effect of $\mathbf{S}_1$ on $\mathbf{H}$ due to unmeasured common factors (causes) is to specify a model $q(t, \bar{h}_m, \bar{s}_{2(m-1)}, \bar{s}_1(t-1), s_{1m}^{\dagger}; \alpha^*)$ for the above contrast, where α^* is an unknown parameter, q is a known function (such as $\alpha[s_{1m} - s_{1m}^{\dagger}]$) that is zero if $s_{1m}^{\dagger} = s_{1m}$ or $\alpha = 0$, so that the magnitude and sign of α^* indicates the degree and direction of confounding due to unmeasured factors. Robins et al. (1999) proved that α^* is non-parametrically unidentified as should be expected since it quantifies the empirically unknowable degree of confounding due to unmeasured factors. Hence, rather than trying to estimate α^* from the data, we regard α^* as known and vary it in a sensitivity analysis as follows. Define the selection bias corrected version $H_t^*(\alpha^*)$ of H_t to be $H_t^*(\alpha^*) = H_t - \sum_{m=0}^{t-1} Q_{m,t}(\alpha^*)$

where $Q_{m,t}(\alpha^*) = \int q(t, \bar{H}_m, \bar{S}_{2(m-1)}, \bar{S}_1(t-1), s^{\dagger}_{1m}; \alpha^*)\, \mathrm{d}F(s^{\dagger}_{1m} \mid \bar{H}_m, \bar{S}_{m-1})$. Robins et al. (1999) proved that the IPTW estimator $\hat{\psi}_{\mathrm{IPTW}} \equiv \hat{\psi}_{\mathrm{IPTW}}(\alpha^*)$ is consistent for the parameter γ^* of the MSM Model (7) when q and α^* are known if, in the definition of $U_{\mathrm{MSM}}(\psi)$, we replace H_t by $H^*_t(\alpha^*)$ and our model for the law of $S_{1t} \mid \bar{H}_t, \bar{S}(t-1)$ is correct. Large sample confidence intervals for $\hat{\psi}_{\mathrm{IPTW}}(\alpha^*)$ can be obtained using standard statistical software. Final conclusions depend on the values of α^* considered plausible. Since the functional form of q is non-parametrically unidentified, several sensitivity analyses using different functional forms should be reported.

Sensitivity analysis for SNMMs: A SNMM for the effect of the final blip of treatment of magnitude s_{1t} among subjects with observed past history $(\bar{h}_t, \bar{s}_{t-1})$ specifies that $\gamma_{tk}(\bar{s}_{1t}, \bar{h}_t, \bar{s}_{2(t-1)}, \psi^*) = \mathrm{E}[H_{d=(\bar{s}_1(t),0),k} \mid s_{1t}, \bar{h}_t, \bar{s}_t] - \mathrm{E}[H_{d=(\bar{s}_1(t-1),0),k} \mid s_{1t}, \bar{h}_t, \bar{s}_t]$, $k > t$ where $\gamma_{tk}(\bar{s}_{1t}, \bar{h}_t, \bar{s}_{2(t-1)}, \psi)$ has the same properties as in Section 6. To conduct a sensitivity analysis, we specify a model $q_{tk}(\bar{h}_t, \bar{s}_{2(t-1)}, \bar{s}_{1t}; \alpha^*)$ for $\mathrm{E}[H_{d=(\bar{s}_1(t-1),0),k} \mid \bar{h}_t, \bar{s}_{2(t-1)}, \bar{s}_{1t}] - \mathrm{E}[H_{d=(\bar{s}_1(t-1),0),k} \mid \bar{h}_t, \bar{s}_{2(t-1)}, \bar{s}_{1(t-1)}, S_{1t} = 0]$ where α^* is a parameter and $q_{tk}(\bar{h}_t, \bar{s}_{2(t-1)}, \bar{s}_{1t}; \alpha)$ is a known function that is zero if $\alpha = 0$ or $s_{1t} = 0$. If $f^{\mathbf{o}}_d[\mathbf{o}] = f^{\mathrm{int}}_d[\mathbf{o}]$ so there is no confounding, then $\alpha^* = 0$. Robins et al. (1999) proved that neither α^* nor q_{tk} is non-parametrically identified. Therefore, we regard α^* as known and vary it in a sensitivity analysis. Define $H^*_{tk}(\alpha^*) = H_k - q_{tk}(\bar{H}_t, \bar{S}_{2(t-1)}, \bar{S}_{1t}; \alpha^*)$. Robins et al. (1999) show that with α^* and q_{tk} known, the g-estimate $\hat{\psi} \equiv \hat{\psi}(\alpha^*)$ is consistent for the SNMM parameter ψ^* if we replace H_k by $H^*_{tk}(\alpha^*)$ in the tk-tera of $U_{\mathrm{SNMM}}(\psi)$.

In the case of MSMMs, an estimate $\hat{\psi}(\alpha^*)$ of ψ^* immediately provides an estimate of $E^{\mathrm{int}}_{d=\mathbf{s}_1}[\mathbf{H}] \equiv \mathrm{E}[\mathbf{H}_{d=\mathbf{s}_1}]$ for any $\mathbf{s}_1$. In the case of SNMMs, knowledge of ψ^*, q and α^* only identifies $E^{\mathrm{int}}_{d=\mathbf{0}}[\mathbf{H}]$. Robins et al. (1999) provide additional non-identifiable assumptions that are sufficient to identify $E^{\mathrm{int}}_d[\mathbf{H}]$ for any $d \in \mathscr{D}^{\mathbf{o}}(\mathbf{S}_1)$.

Robins et al. (1999) also derived variation independent parameterizations for the joint distribution of the observed data $(\bar{S}_1, \bar{S}_2, \bar{H})$ compatible with a given MSMM or given SNMM and their respective selection bias functions q and sensitivity parameters α^*. These parameterizations allow Bayesian or parametric likelihood based estimation of ψ^* under either model. Because of lack of identifiability when q and $\alpha^*$are unknown, it will be essential for decision-making purposes to augment any sensitivity analysis with a Bayesian analysis in which plausible priors are specified for the selection bias functions q and sensitivity parameters α^* (jointly with the other parameters in the likelihood).

A Bayesian perspective also allows us to precisely characterize the relationship of a faithfulness analysis of causation to a sensitivity analysis approach based on MSMs or SNMs in the context of Example 4. Specifically suppose, that as in Example 4, the following are satisfied by the true but unknown law of $\mathbf{O}$: Eq. (1) is false and the "g"-null hypothesis based on data $\mathbf{O}$ is true for the effect $\mathbf{S}_1$ on $\mathbf{H}$ but not on $\mathbf{S}_2$. Then the calculations in Robins et al. (1999, Chapter 11) can be generalized to show that, for a panel data study with sample size n, if the prior for α^* puts positive mass p on 0 (i.e., on the hypothesis of no unmeasured confounders for the effect of $\mathbf{S}_1$ on $\mathbf{H}$ given data on $\mathbf{O}$) then, as $n \to \infty$, the posterior distribution for α^* will converge to a point mass at 0 and the posterior for the causal parameter ψ^* for a SNMM and ψ^*_1 for a MSMM will converge to a point mass at their null value of 0. That is, our inferences will agree precisely with those obtained by a faithfulness analysis: in large samples, we will conclude both no unmeasured confounding and no causal effect of $\mathbf{S}_1$ on $\mathbf{H}$. However, if we use a smooth prior density for α^* with respect to Lesbesgue measure so that no point, including 0, is assigned a positive prior probability, then even as $n \to \infty$, our posteriors for α^* and ψ^* of a SNMM (ψ_1^* of a MSMM) will not converge to a point but rather will be distributions whose location and spread will depend quite sensitively on the joint prior for the model. Thus, if, as I recommend, we do not assign a positive probability to the event of no unmeasured confounders ($\alpha^* = 0$), our causal inferences will depend strongly on the precise form of our prior beliefs, regardless of the sample size.

References

Cox, D.R., Wermuth, N., 1999. Likelihood factorizations for mixed discrete and continuous variables. Scandinavian Journal of Statistics 25, 209–220.

Greenland, S., Robins, J.M., 1986. Identifiability, exchangeability and epidemiologic confounding. International Journal of Epidemiology 15, 413–419.

Humphreys, P., Freedman, D., 1996. The grand leap. British Journal for the Philosophy of Science 47 (1), 113–123.

Lauritzen, S.L., Dawid, A.P., Larsen, B.N., Leimar, H.G., 1990. Independence properties of directed Markov fields. Networks 20, 491–505.

Pearl, J., 1995. Causal diagrams for empirical research. Biometrika 82 (4), 669–688.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, Los Atlos, CA.

Pearl, J., 2000. Causality. Cambridge University Press, Cambridge, UK.

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

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 (9–12), 923–945.

Robins, J.M., 1989. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest, L, Freeman, H, Mulley, A. (Eds.), Health Service Research Methodology: A Focus on AIDS. U.S. Public Health Service, National Center for Health Services Research, Washington, D.C. pp. 113–159.

Robins, J.M., 1994. Correcting for non-compliance in randomized trials using structural nested mean models. Communications in Statistics 23 (8), 2379–2412.

Robins, J.M., 1995. Discussion of "Causal Diagrams for Empirical Research" by J. Pearl. Biometrika. Biometrika 82 (4), 695–698.

Robins, J.M., 1997. Causal inference from complex longitudinal data. In: Berkane, M (Ed.), Latent Variable Modeling and Applications to Causality, Lecture Notes in Statistics, Vol. 120. Springer, New York, pp. 69–117.

Robins, J.M., 1999. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran, M.E., Berry, D. (Eds.), Statistical Models in Epidemiology: The Environment and Clinical Trials. Springer, New York, pp. 95–134.

Robins, J.M., 2002. Semantics of causal DAG models and the identification of direct and indirect effects. In: Green, P., Hjort, N., Richardson, S. (Eds.), Highly Structured Stochastic Systems. Oxford University Press, Oxford, to appear.

Robins, J.M., Wasserman, L., 1997. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In: Geiger, D., Shenoy, P. (Eds.), Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence, Providence, RI, August 1–3, 1997. Morgan Kaufmann, San Francisco, pp. 409–420.

Robins, J.M., Wasserman, L., 1999. On the impossibility of inferring causation from association without background knowledge. In: Glymour, C., Cooper, G. (Eds.), Computation, Causation, and Discovery. AAAI Press/The MIT Press, Menlo Park, CA, Cambridge, MA, pp. 305–321.

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

Robins, J.M., Scharfstein, D., Rotnitzky, A., 1999. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran, M.E., Berry, D. (Eds.), Statistical Models in Epidemiology: The Environment and Clinical Trials. Springer, New York, pp. 1–94.

Spirtes, P., Glymour, C., Scheines, R., 1993. Causation, Prediction, and Search. Springer, New York.