Statistical Science
2014, Vol. 29, No. 3, 367–370
DOI: 10.1214/14-STS488
Main article DOI: 10.1214/14-STS480

Causal Graphs: Addressing the Confounding Problem Without Instruments or Ignorability

Ilya Shpitser

1. INTRODUCTION

I wish to congratulate Professor Imbens on a lucid and erudite review of the instrumental variable literature. The paper contrasts an econometric view of instrumental variable models, where treatment confounding is due to agents rationally choosing an optimal treatment for their situation, and the statistical view, where treatment confounding arises due to noncompliance, unobserved baseline differences between individuals, or other such issues.

While the paper does an admirable job describing the statistics view of the instrumental variables based on the potential outcome model of Neyman and Rubin, it does not much discuss the growing statistics literature on causal graphical models, except to mention that causal graphs are a useful tool for displaying the exclusion restriction assumption crucial for the use of instrumental variables.

I would like to give a brief and hopefully complementary account of how causal graphical models serve to clarify and help address the issues of confounding (what Heckman calls the selection problem) that make causal inference from observational data such a challenging endeavor.

2. GRAPHS AS A GENERAL METHOD FOR DEALING WITH CONFOUNDING

Causal inference in statistics has been greatly influenced by Neyman's idea of explicitly representing interventions or forced treatment assignments on the outcome (Neyman, 1923), and by Rubin's idea of using the stable unit treatment value assumption (SUTVA) and ignorability assumptions to equate potential outcome parameters with functionals of the observed data (Rubin, 1974). Professor Imbens discusses these ideas at length in the paper. The essence of Rubin's method is that assumptions on potential outcome random variables allow one to properly adjust for the presence of confounding. Unfortunately, in complex, possibly longitudinal settings it is not easy to see what assumptions are needed, or whether it is even possible to identify parameters of interest as functionals of observed data. For this task, graphical causal models, first used by Wright in the context of animal genetics (Wright, 1921), and expanded into a general methodology for causal inference by Spirtes, Glymour and Scheines (1993), Pearl (2000), Robins (1986, 1997), and others have proven to be invaluable.

Consider Figure 1(a), where vertices represent random variables of interest: a treatment A, an outcome Y, and a source of unobserved confounding C (lightly shaded in the graph to represented unobservability). Following Neyman, we quantify the causal effect of A on Y by means of a function of the distribution of the potential outcome $Y(a)$ (Y after we force A to a value a). For instance, we may use the average causal effect (ACE): $E[Y(a)] - E[Y(a')]$, where a is the active treatment value, and a' is the baseline treatment value. We are interested in using observed data to make inferences about such effects, which entails dealing with confounding in some way. The assumptions underlying this graph which we will use can be expressed in terms of potential outcomes if desired. For example, the *finest fully randomized causally interpretable structured tree graph* (FFRCISTG) model of Robins (1986) corresponding to this graph states that for all value assignments a and c to A and C, random variables C, $A(c)$ and $Y(a, c)$ are mutually independent, while the *nonparametric structural equation model with independent errors* (NPSEM-IE) of Pearl (2000) corresponding to this graph states that for all value assignments a, c and c' to A and C, random variables C, $A(c)$ and $Y(c', a)$ are mutually independent. Note that the former set of assumptions can be viewed as a kind of

Ilya Shpitser is Lecturer in Statistics, Department of Mathematics, University of Southampton, University Road, Building 54, Southampton, Hampshire, SO17 1BJ, United Kingdom (e-mail: i.shpitser@soton.ac.uk).

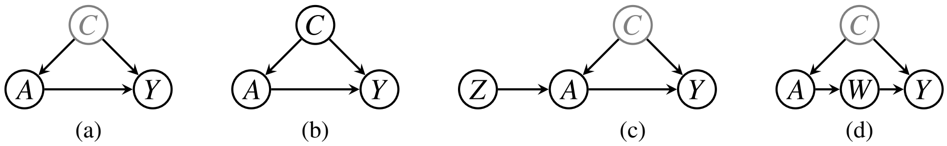

FIG. 1. (a) *The standard problem of causal inference—an unobserved confounder C, and possible approaches to the problem using observational data*. (b) *Observing the confounder and adjustment/stratification methods*. (c) *Z as an instrumental variable*. (d) *A strong independent mediator as an "instrument" for identification*.

mutual ignorability assumption derived from the graph, while the latter set can be viewed as a mutual version of what Imai called the *sequential ignorability* assumption (Imai, Keele and Yamamoto, 2010). A general method for associating an arbitrary graph with sets of assumptions on potential outcomes can be found, for instance, in the paper by Richardson and Robins (2013). Thus, graphs are merely a visual representation of familiar potential outcome models. A purely visual view, however, can prove quite helpful.

A common approach within the Rubin framework is to assume that a conditional ignorability assumption ($Y(a) \perp\!\!\!\perp A \mid C$) holds. Here, $\perp\!\!\!\perp$ is the conditional independence symbol. This assumption logically follows from the assumptions defining the FFRCISTG model of Figure 1(a).[1] Moreover, if C is observed (represented graphically by Figure 1(b), where C is now normally shaded), this assumption in turn entails that the distribution of $Y(a)$ can be expressed as a functional of the observed data via the adjustment formula

$$p(Y(a)) = \sum_c p(Y \mid a, c) p(c)$$

which in turn can be estimated by a variety of methods, including propensity score methods (Rosenbaum and Rubin, 1983), inverse weighting methods (Horvitz and Thompson, 1952), the parametric g-formula (Robins, 1987) or doubly robust methods (Robins, Rotnitzky and Zhao, 1994).

If conditional ignorability is not a sensible assumption, or we cannot make use of it due to strong sources of confounding that cannot be measured, as is often the case in econometric applications, we may instead try to find an instrument, which is shown graphically in Figure 1(c). Here, the missing arrow from Z to Y represents the exclusion restriction assumption namely that $Y(a, z) = Y(a, z')$ for any a, z and z', and the lack of arrows from C to Z represents the assumption that the instrument behaves as if randomly assigned: $Z \perp\!\!\!\perp \{A(z), Y(a, z)\}$ for any a, z. These assumptions also logically follow from the assumptions defining the FFRCISTG model for the graph in Figure 1(c). As Professor Imbens discusses, given these assumptions, one could obtain bounds on the effect, or using further assumptions, obtain point identification.

It may be that we cannot make sure of the conditional ignorability assumption (that is, large parts of C are not observable), and it is the case that a variable for which the above instrumental assumptions hold cannot be found. In this case, it is possible to use the systematic representation of restrictions on potential outcomes given by a graph to derive additional methods of attack on the confounding problem which can complement the instrumental variable and conditional ignorability approaches.

For example, it may be possible that a strong *mediating* variable for the effect of A on Y exists and is observable, and moreover, the mechanism by which this mediation happens is independent of the source of the confounding between the treatment A and outcome Y, given the treatment A. This situation is shown in Figure 1(d). In terms of assumptions on potential outcomes, such a strong mediator between A and Y is represented as stating that $Y(w, a) = Y(w, a')$ for all w, a, a'. In words, A has no direct effect on Y once we fix W to any value w. The unconfoundedness of the mediator may be represented as stating that $\{Y(w), A\} \perp\!\!\!\perp W(a)$. It is easy to see that the former assumption is a kind of exclusion restriction corresponding to the absence of a directed arc from A to Y, and the latter is a kind of ignorability assumption, which corresponds to the absence of an arc from C to W, and the absence of other sources of confounding between W and A and Y. Professor Imbens discusses these kinds of assumptions in the context of instrumental variables in Sections 5.1 and 5.2. Pearl has shown a result that is equivalent to stating that given a version of SUTVA,

[1]The NPSEM-IE always makes at least as many assumptions as the FFRCISTG model, and in many cases more. Thus, any assumption entailed by the FFRCISTG model of a graph is also entailed by the NPSEM-IE for the same graph.

and above assumptions, the following *front-door formula* holds:

$$p\big(Y(a)\big) = \sum_{w} p(w \mid a) \sum_{a'} p\big(Y \mid w, a'\big) p\big(a'\big). \quad (2.1)$$

As before, the above assumptions logically follow from the assumptions defining the FFRCISTG model for Figure 1(d).

A (slightly contrived) example of a situation represented in Figure 1(d) is as follows. We may suspect that consumption of skyr (a kind of Icelandic dairy product) is protective against stomach cancer by means of a particular type of flora found in skyr. However, we suspect those with Icelandic citizenship may both consume more skyr than average, and have different rates of stomach cancer than the general population. If we cannot observe citizenship status, but we can do a simple test for the presence of the protective flora, and moreover, we suspect the protective causal mechanism is not influenced by the confounding variable directly but only through skyr consumption, and moreover, this mechanism mediates *all* of the effect of skyr consumption, then we can find a way to express the causal effect of skyr consumption on incidence of stomach cancer using observational data via (2.1).

Since graphs are merely a systematic *visual* way of arranging information on potential outcomes, they have proven extremely helpful for generating solutions to problems posed by confounding in very general settings. For instance, VanderWeele and Shpitser (2013) used graphs to show that many informal definitions for what a "confounder" is in the literature are "incorrect" in the sense of not agreeing with intuition in given examples, and not obeying certain natural properties we expect a confounder to obey, while a definition that is "correct" in this sense is fairly subtle.

Furthermore, a mediator-based approach resulting in (2.1) has been generalized using causal graphs to a fully general method of "deconfounding" which can successfully be applied in complex longitudinal settings even when no standard ignorability assumptions can be used, and no good instruments can be found.

Consider a hypothetical longitudinal study represented by the causal graph shown in Figure 2, where bidirected arrows represent the presence of some hidden common cause. For instance, a bidirected arrow from A to C means there is a hidden common cause of A and C. In this study, B and D are administered treatments, Y is the outcome, A is an observed baseline confounder, and C is an intermediate health measure. These variables are confounded, but in a very particular way displayed by bidirected arrows in the graph.

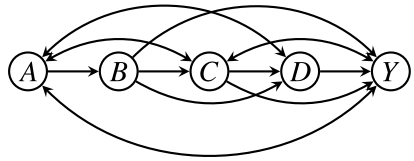

FIG. 2. *A longitudinal study with exposures B, D, outcome Y and multiple sources of unobserved confounding.*

For instance, there is no hidden common cause of B and any other variable, and D has a hidden cause in common only with A.

We are interested in the total effect of the drug on the outcome, which can be obtained from the distribution of $Y(b, d)$, in order to better determine the optimal treatment assignment. Robins (1986) gave a general method for expressing $Y(b, d)$ as a functional of the observed data given that an assumption called sequential ignorability (different from one in Imai, Keele and Yamamoto, 2010) holds. Sequential ignorability on the observable variables happens not to be implied by either the FFRCISTG model, or the NPSEM-IE of any underlying hidden variable DAG consistent with Figure 2, but it can be shown that assumptions implied by these models can be used to derive the following identity:

$$p\big(Y(b,d) = y\big) = \sum_{c} \Big(\sum_{a} p(y, d, c \mid b, a) p(a) \Big) \cdot \frac{\sum_a p(c \mid b, a) p(a)}{\sum_a p(d, c \mid b, a) p(a)}. \quad (2.2)$$

Just as in the previous cases, it is possible to derive in a systematic way a list of restrictions on potential outcomes over observable variables implied by the graph in Figure 2, and use this list to derive (2.2) without referring to the graph at all. However, without the help of the graph it is not so easy to see what this list of restrictions might look like, or how the derivation based on this list might proceed. In more complex longitudinal settings, with many more than 5 variables the problem becomes even more severe.

A general method for identifying potential outcome distributions as functionals of observed data for causal graphs was given in the paper by Tian and Pearl (2002), and was proven complete (in the sense of only failing on nonidentifiable distributions) by Shpitser and Pearl (2006a, 2006b, 2008), Huang and Valtorta (2006). While it is possible to rederive these results in terms of assumptions on potential outcomes, graph theory provided mathematical terminology and intuitions that proved crucial in practice for deriving these general "deconfounding" results.

3. CONCLUSION

Professor Imbens' excellent review shows that in the context of economics, confounding arises due to the agent's decision algorithm, while in the context of statistics confounding arises for other reasons (for instance, due to lack of compatibility among patients in an observational study). Instrumental variables are an important technique for dealing with confounding in causal inference, though alternative methods involving stratification were developed under the potential outcome model of Rubin and Neyman.

More general lines of attack on the problem of confounding were developed using graphical causal models, first used by Wright in genetics, and extended into a general model by Spirtes, Glymour and Scheines (1993), Pearl (2000), Robins (1986, 1997) and others. These methods allow "deconfounding" of the problem even in cases where confounders cannot be observed, and good instrumental variables do not exist.

REFERENCES

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

HUANG, Y. and VALTORTA, M. (2006). Pearl's calculus of interventions is complete. In *Proceedings of the* 22*nd Conference on Uncertainty in Artificial Intelligence* 217–224. AUAI Press, Arlington, VA.

IMAI, K., KEELE, L. and YAMAMOTO, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statist. Sci.* **25** 51–71. MR2741814

NEYMAN, J. (1923). Sur les applications de la thar des probabilities aux experiences Agaricales: Essay des principle. Excerpts reprinted in English: *Statist. Sci.* **5** (1990) 463–472.

PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge Univ. Press, Cambridge. MR1744773

RICHARDSON, T. S. and ROBINS, J. M. (2013). Single World Intervention Graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Preprint, available at http://www.csss.washington.edu/Papers/wp128.pdf.

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—Application to control of the healthy worker survivor effect. *Math. Modelling* **7** 1393–1512. MR0877758

ROBINS, J. M. (1987). A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Diseases* **40** 139–161.

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

ROBINS, J. M. and WASSERMAN, L. (1997). Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In *Proceedings of the* 13*th Conference on Uncertainty in Artificial Intelligence* 409–420. Morgan Kaufmann, San Francisco, CA.

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology* **66** 688–701.

SHPITSER, I. and PEARL, J. (2006a). Identification of conditional interventional distributions. In *Proceedings of the* 22*nd Conference on Uncertainty in Artificial Intelligence* 437–444. AUAI Press, Arlington, VA.

SHPITSER, I. and PEARL, J. (2006b). Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence* 1219–1226. AAAI Press, Menlo Park, CA.

SHPITSER, I. and PEARL, J. (2008). Complete identification methods for the causal hierarchy. *J. Mach. Learn. Res.* **9** 1941–1979. MR2447308

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (1993). *Causation, Prediction, and Search. Lecture Notes in Statistics* **81**. Springer, New York. MR1227558

TIAN, J. and PEARL, J. (2002). On the testable implications of causal models with hidden variables. In *Proceedings of the* 18*th Conference on Uncertainty in Artificial Intelligence* 519–527. Morgan Kaufmann, San Francisco, CA.

VANDERWEELE, T. J. and SHPITSER, I. (2013). On the definition of a confounder. *Ann. Statist.* **41** 196–220. MR3059415

WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research* **20** 557–585.