

Strategies of Research Design with Confounding: A Graphical Description*

Adam N. Glynn[†] and John Gerring[‡]

May 25, 2013

Abstract

Research design is of paramount importance when attempting to overcome confounding. In this paper, we propose a unified graphical approach for the consideration of cross-sectional research designs. Specifically, we argue that at least five distinct strategies may be discerned for coping with the presence of a common-cause confounder: (1) blocking backdoor paths, (2) mechanisms, (3) instrumental variables, (4) alternate outcomes, and (5) causal heterogeneity. All of these strategies enlist a facilitating variable, whose role defines the corresponding research design. This resulting framework builds on the foundational work of Pearl (2000, 2009) but incorporates additional research designs into the graphical framework, providing a more comprehensive typology of designs for causal inference.

*Preliminary draft produced for the 2011 IQMR at Syracuse University. Please do not cite or circulate without permission of the authors.

[†]Department of Government, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138. aglynn@fas.harvard.edu

[‡]Department of Political Science, Boston University, 232 Bay State Road, Boston MA 02215. jgerring@bu.edu

1 Introduction

Observational data is prone to common-cause confounders (C), i.e., factors that affect both the theoretical variable of interest (X) and the outcome of interest (Y), and whose presence renders the covariation between X and Y spurious. This may be labeled an assignment problem, a selection effect, or non-ignorability. All refer to the fact that the potential outcomes are not independent of assignment to treatment, a problem lying at the heart of causal inference with observational data. Indeed, the great variety of research designs and econometric strategies that social scientists and statisticians have developed are, in large part, attempts to deal with threats emanating from common-cause confounders.

With the assistance of a causal-graph framework, we identify five strategies for achieving identification in the presence of a common-cause confounder: (1) blocking backdoor paths, (2) mechanisms, (3) instrumental variables, (4) alternate outcomes, and (5) causal heterogeneity. All of these strategies enlist a facilitating variable, labeled Z in the following discussion. Likewise, the specific role of Z defines the distinguishing characteristic of each research design, as illustrated in graphs that accompany the text. Solving problems of causal inference in the presence of a common-cause confounder can thus be thought of as an attempt to identify additional variables that, in one way or another, overcome problems of inference where conditioning on X and Y alone will not suffice. We note at the outset that these strategies are not equally promising, but the following discussion will reveal the pros and cons of each strategy.

This strategy builds on Pearl (2000, 2009) and Morgan and Winship (2007: 26), who focus on the first three strategies. We provide a more comprehensive typology by presenting additional graphical representations of some of the strategies considered in Angrist and Pischke (2008), and the “devices” presented in Rosenbaum (2002, 2009),

as well as some minor variations. In this manner, we hope to identify a capacious toolkit of methods for causal inference, while clarifying the similarities/differences and strengths/weaknesses of each.

Causal graphs have several attractive properties as a framework for thinking about research design. First, with minimal notation they allow us to formalize, and thereby unify, the problem of confounders and methods for dealing with them across diverse research traditions. It is our contention that many cross-sectional research designs - including those employing quantitative and qualitative methods - can be subsumed within the foregoing typology. Second, causal graphs provide a link between a presumed data generating process and econometric techniques that might be employed for estimating causal effects. Properly deployed, they embed statistics in social reality. Third, the framework is highly intuitive. A visual representation of key research design elements often helps to clarify the varying logics of causal inference and may assist in the practical task of identifying appropriate research designs to suit various settings. Finally, causal graphs provide a mathematically rigorous framework consistent with nonparametric structural equations models and potential outcomes models (Pearl, 1998, 2000, 2009; Spirtes et al. 1998, 2000). Granted, some of the designs require the addition of parametric information to the nonparametric graphs in order to explain the assumptions utilized in a particular research design. This suggests a further utility of the five-part typology: designs differ in their structure (denoted by the graph associated with each design) and in the assumptions required for the identification of average effects.

The paper begins by introducing the problem of confounders and the framework of causal graphs. In succeeding sections we present the five research design strategies. For each strategy, we provide a graph (or set of graphs), an example (or several), and a discussion of key assumptions. A short conclusion adds some important caveats and

discusses the import of a graphical approach to causal inference.

2 Confounders through the Lens of Causal Graphs

We begin with a brief review of directed acyclic graphs (DAGs), as developed by Judea Pearl (2009). In this framework, the variables in a model are represented as nodes in a graph. Each node may represent a single variable or a vector. We shall speak of variables rather than vectors in order to simplify the exposition; however, readers should bear in mind that a node may represent a vector if a set of variables bears a similar relationship to other nodes.

The key nodes in any DAG are X and Y , exemplifying the factors of theoretical interest. Other factors may be included if they serve a role in causal identification. Other factors *must* be included if they impact more than one node of the graph. An example of this is the common-cause confounder, which by definition affects X and Y and thus must be included in a DAG if it is thought to exist. A DAG with a confounder (C) is illustrated in Figure 1.

Variables represented in a DAG may be measured and potentially conditioned (signaled by a solid circle) or not (signaled by an open circle). This evidently makes a great deal of difference in any analysis of causal effects, so the reader will want to keep a close eye on the way in which nodes are represented in the following graphs. In Figure 1, the confounder is not measured. Under the circumstances, the resulting relationship between X and Y may be spurious: a correlation between X and Y might be the product of (a) a causal relationship or (b) the unmeasured and unconditioned factor, C , which impacts both X and Y .

Whenever one variable is suspected of affecting another, an arrow is drawn between the two nodes. Missing arrows correspond to exclusion restrictions, i.e., the presumed lack of a direct causal effect. A DAG is “acyclic” in that arrows point in only one

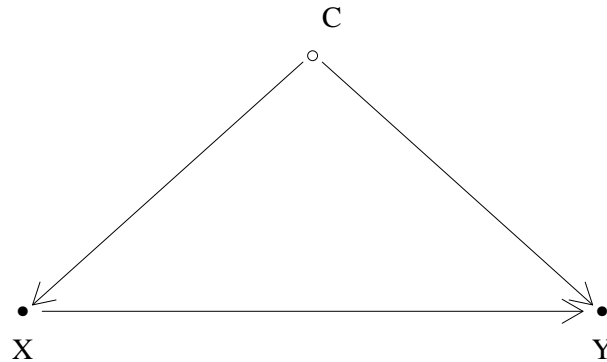


Figure 1: *Directed Acyclic Graph (DAG)*. Open circles correspond to unmeasured (and therefore unconditioned) variables. Solid circles correspond to measured variables. Missing arrows correspond to exclusion restrictions (the lack of a direct causal effect) and all common causes are represented on the graph.

direction. Likewise, there is no way to start at one variable and return to that same variable by following the arrows on the graph.

In order to see how this works it is necessary to familiarize oneself with several rules of conditioning. Consider the four panels in Figure 2. In panel (a), we expect an association between X and Y because X causes Y . In panel (b), after conditioning on Z , we expect no association between X and Y because we have conditioned on Z , an intermediary variable, which thereby breaks the flow of information from X to Y (or Y to X). In panel (c), after conditioning on Z , we expect no association between X and Y because we have conditioned on Z , a common cause of both X and Y . In panel (d), we expect no unconditional association between X and Y but we do expect a conditional association between X and Y because we have conditioned on Z , a factor that is affected by both X and Y . This setting, where Z serves as a “collider,” is somewhat counterintuitive and therefore bears special emphasis: a relationship can be created between two orthogonal variables by conditioning on a third variable that both

affect. The full set of rules for reading conditional independence relationships from a graph is known as d-separation (Pearl 2000, pg. 16).

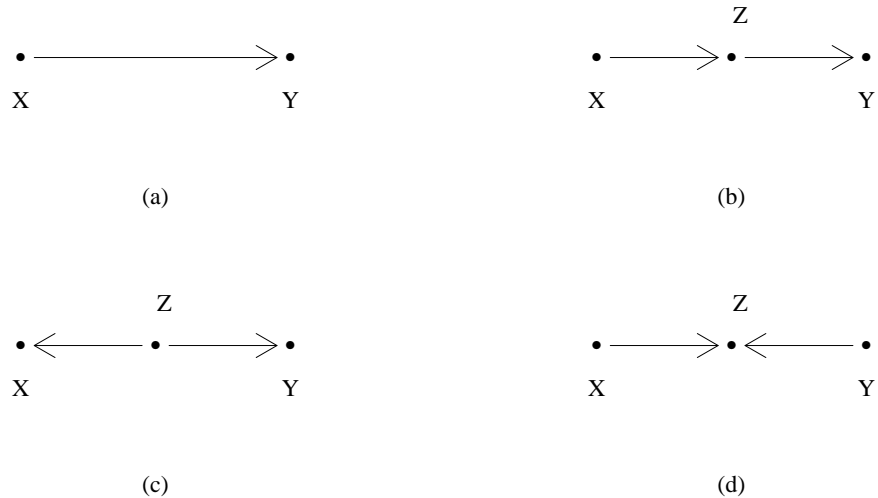


Figure 2: *Basic Principles of Conditional Association Illustrated for DAGs.* In panel (a), we expect an association between X and Y because X causes Y . In panel (b), after conditioning on Z , we expect no association between X and Y because we have conditioned on Z , an intermediary variable, which thereby breaks the flow of information from X to Y (or Y to X). In panel (c), after conditioning on Z , we expect no association between X and Y because we have conditioned on Z , a common cause of both X and Y . In panel (d), we expect no unconditional association between X and Y but we do expect a conditional association between X and Y because we have conditioned on Z , a factor that is affected by both X and Y .

For heuristic purposes it is helpful to develop a single research question that can be applied to numerous empirical settings, thus unifying our discussion of diverse approaches to causal inference. Our exemplar is an imaginary after-school ESL (English as second language) program established as a supplement for secondary school students in the United States who are not native speakers. The research question of theoretical interest is whether the program has a causal impact on English-language capacities, all other things being equal. Impact refers to average treatment effects (ATE), unless otherwise noted in the following discussion. Importantly, the program is free and sign-up is voluntary and thus open to potential confounders: non-native speakers who

sign-up may be different from those who do not, in ways that affect their acquisition of English. (In this respect, our chosen exemplar is similar to other settings that have received extensive commentary in recent years such as job-training programs [Heckman et al. 1998; Lalonde 1986].) Specifically, those who attend may be more motivated or better prepared, and this in turn may be the product of a prior factor: the educational attainment of their parents.

Returning to Figure 1, X represents the treatment (attendance in the ESL program), Y represents the outcome (English proficiency, as measured by language tests), and C represents various possible confounders (as listed above). Of course, this figure, like all DAGs, encodes a set of assumptions, which in the event may or may not be true. In this respect, it is no different from other models of causal inference - formal, statistical, or expository. Nonetheless, insofar as all causal inference rests on assumptions it is important to be explicit about what these are, and the properties of the DAG are enormously helpful in clarifying some (but not all) of the necessary assumptions.

3 Blocking Backdoor Paths

The usual approach to dealing with a common-cause confounder involves blocking “backdoor” paths. Specifically, if there are no causal cycles, and if all values of the causal factor X are possible for all values of the conditioning variables (e.g., at least some children of both high and low education parents must have signed up for the ESL program), then the average effect of X on Y can be identified when (a) the set of conditioning variables does not include variables affected by X , and (b) the set of conditioning variables blocks all paths between X and Y that point back into X (i.e., the backdoor paths of spurious correlation/confounding). These two conditions are known as the backdoor criterion (Pearl 2000, pg. 79).

Returning to our example, let us stipulate that the key difference between partic-

ipants and non-participants in the after-school ESL program is that the former are more likely to have parents who are well educated. These college-educated parents sign their children up for the after-school program (affecting assignment to treatment) and help motivate their children to work on the assignments (affecting the outcome of achievement tests), thus serving as a common-cause confounder.

One approach to blocking backdoor paths involves conditioning on the confounder itself. Thus, one might measure the educational status of parents, including this as a covariate in the analysis. As illustrated in panel (a) of Figure 3, we write this conditioning variable as Z (rather than C) in order to emphasize its role in the causal analysis. This is the first example of what we will refer to as a facilitating variable - a variable of no theoretical interest that facilitates causal inference.

Another approach is to block the path from C to Y , i.e., the effect of the confounder on the outcome, as illustrated in panel (b) of Figure 3. Returning to our example, one can imagine measuring the impact of parents on student work efforts, e.g., by asking students whether their parents helped to motivate them to complete class assignments, and then including the resulting variable as a covariate in the analysis.

A third approach is to block the path from C to X , i.e., the effect of the confounder on assignment to treatment, as illustrated in panel (c) of Figure 3. This might be accomplished in the present setting by measuring the selection process, i.e., the mechanism by which students decided to sign up for the after-school class. Specifically, one might find a way to distinguish those who signed up of their own accord from those who signed up due to parental pressure, including the resulting variable as a covariate in the analysis.

It will be seen that blocking backdoor paths does not necessarily entail conditioning on the confounder. Equally important, although not the focus of this article, conditioning on "irrelevant variables" (i.e., variables that neither affect nor are affected by X

or Y) can lead to spurious correlation when these variables are colliders (as in Figure 2 (d)) on a backdoor path. Depending on the definition of irrelevant, this result is contrary to advice in many econometrics textbooks.¹

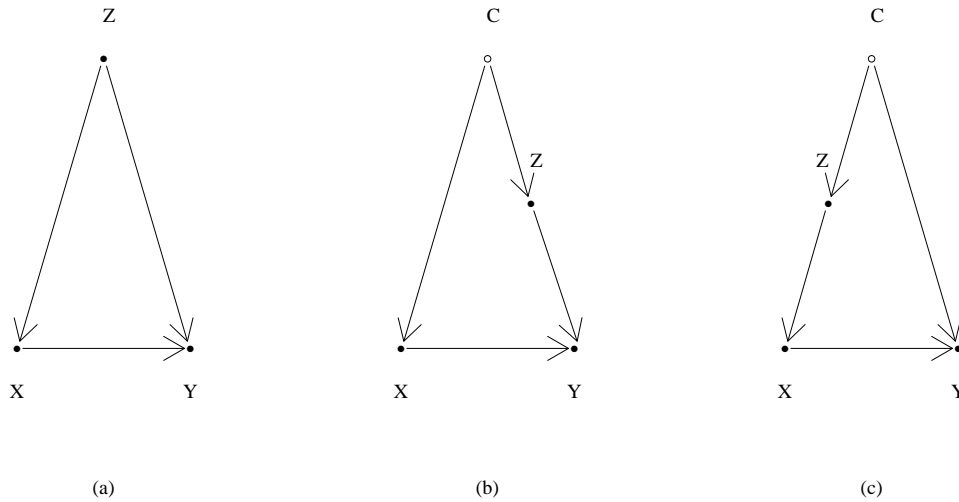


Figure 3: *Blocking Backdoor Paths*. In panel (a), the backdoor path is blocked by conditioning on the common cause. In panel (b), the backdoor path is blocked by conditioning on a variable that blocks the path from the confounder to the outcome (Y). In panel (c), the backdoor path is blocked by conditioning on a variable that blocks the path from the confounder to the causal variable of interest (X).

4 Mechanisms

If we cannot find variables to block backdoor paths, a second approach to confounding uses causal mechanisms, understood as the variables on connecting path(s) between X and Y , as facilitating variables. This approach has been invoked in a number of forms by qualitative and quantitative researchers and is referred to variously as process tracing, process analysis, causal narrative, colligation, congruence, contiguity, discerning, intermediate processes, or microfoundations.²

¹See Elwert and Winship (2012) for examples.

²Terms and associated works are as follows: process-tracing (George and McKeown, 1985, pg. 34ff), discerning (Komarovsky, 1940, pg.135-46), process analysis (Barton and Lazarsfeld, 1969), pattern-matching (Campbell, 1975), microfoundations (Little, 1998), causal narrative (Abbott, 1996a,b; Abrams, 1982; Am-

Pearl’s contribution to this literature is the demonstration that mechanisms can be used to nonparametrically identify the average effect of X on Y , even when confounding cannot be blocked. Pearl (2000, 2009) demonstrates that a set of mediating variables (Z) can be used to identify the average effect of X on Y when (a) Z intercepts all directed paths from X to Y , (b) there is no path from X to Z with an arrow into X that is not blocked, and (c) there is no path from Z to Y with an arrow into Z that is not blocked by X (Pearl 2000, pg. 82). This is known as the front-door criterion. Tian and Pearl (2002) provides a generalization and Knight and Winship (2012) provides some additional extensions.

Figure 4 (a) presents a DAG where the front-door criterion holds. Although the open circle at C indicates that this variable is unmeasured, and therefore there is no conditioning set that will satisfy the backdoor criterion for the average effect of X on Y , notice that the variable Z in this graph satisfies the front-door criterion. There is only one directed path from X to Y and Z intercepts this path. The only path from X to Z with an arrow into X goes through the collider at Y , so it is blocked by the empty set. This implies that the backdoor criterion is satisfied for the effect of X on Z . Finally, the only path from Z to Y with an arrow into Z goes through the chain at X , so it is blocked by X . This implies that the backdoor criterion is satisfied for the effect of Z on Y by conditioning on the variable X . Because the backdoor criterion is satisfied for both the effect of X on Z and the effect of Z on Y , we can recompose the average effect of X on Y by combining the two component effects (Pearl 2000, pg. 83).

To date, there have been relatively few applications of the front-door criterion (although see Winship and Harding (2008)), and it remains a largely hypothetical research

inzade, 1992; Bates et al., 1998; Griffin, 1992, 1993; Katznelson, 1997; Kiser, 1996; Mink, 1987; Quadagno and Knapp, 1992; Roth, 1994; Rueschemeyer and Stephens, 1997; Sewell, 1992, 1996; Somers, 1992; Stone, 1979; Stryker, 1996; Watkins, 1994), congruence (George and Bennett, 2005), colligation (Roberts, 1996), intermediate processes (Mill, 1843/1872). For general discussion see Bennett (1999), (Brown, 1984, pg.228), (Collier and Mahoney, 1996, pg.70), Goldstone (1997). For philosophical discussion in the realist tradition, see Bhaskar (1975/1978); Harre (1972); McMullin (1984); Salmon (1984).

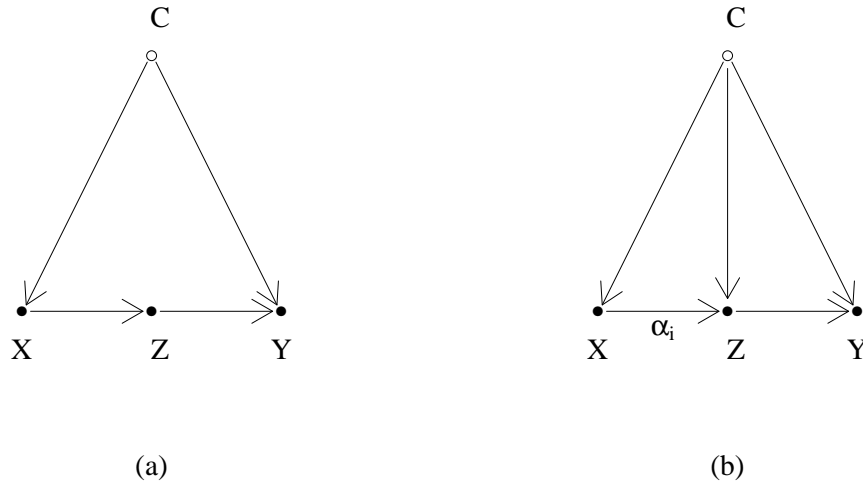


Figure 4: *DAGs associated with the strategy of mechanisms* In panel (a), the front-door criterion holds. Z intercepts all directed paths from X to Y , there is no path from X to Z with an arrow into X that is not blocked by the empty set, and there is no path from Z to Y with an arrow into Z that is not blocked by X . In panel (b), the front-door criterion does not hold, but we assume a linear heterogeneous effect of X on Z (denoted by the inclusion of α_i on the graph). If X is binary then linearity holds trivially, but if we further assume that $\alpha_i \geq 0$ and $Z_i \geq 0$ for all i , causal effects can be ruled out for all i with $X_i = 1$ and $Z_i = 0$.

design (Vanderweele, 2009). However, there are alternative mechanistic approaches to ruling out causal effects that follow some aspects of the front-door approach. Specifically, the assumption of a missing arrow from C to Z is replaced with assumptions regarding the possible values of X , Z and the effect of X on Z . Often this takes the form of binary X and $Z \geq 0$ with a monotonicity assumption— that X can't have negative effects on Z for any individual. If we define the effect of X on Z in terms of potential outcomes,

$$Z(x) = \epsilon_i + \alpha_i x$$

then the monotonicity assumption implies that $\alpha_i \geq 0$ for all i . As with path diagrams of linear structural equations models (linear SEMs), when we are assuming linear effects, we will include the parameters associated with these effects on the arrows of the

graph (see Figure 4 (b)). However, note that the inclusion of i subscripts denotes the possibility of effect heterogeneity.

In our running example, such a mechanistic approach might take the form of measuring class attendance as Z . If we are willing to assume that the ESL program only has an effect through attendance (i.e., the missing arrow from X to Y), and we are willing to assume that you cannot attend without signing up, then monotonicity holds and we can assume that the program has no effect for those that sign up but do not attend. If the outcome variable Y is bounded (e.g., pass/fail), this approach can provide an upper bound on the average treatment effect among those that actually signed up for the program, sometimes known as Average Treatment effect on the Treated (ATT).³

5 Instrumental Variables

One of the most familiar research designs for dealing with an unmeasured confounder uses a facilitating variable known as an instrument. Structurally, this strategy is somewhat similar to the front-door approach of the previous section in that the facilitating variable Z must be unaffected by the confounder. However, unlike a mediating variable, an instrument is antecedent to X . This strategy also requires an assumption that Z has no direct effect on Y that does not go through X .

The DAG associated with this strategy is presented in Figure 5 (a), and the structure of the graph provides some idea of how the instrumental variables (IV) design provides information about the effects of X on Y . The structure implies that the back-door criterion holds for the effect of Z on X and also for the effect of Z on Y . This result is straightforward for the effect of Z on X because the arrow from X to Z is the only path. For the effect of Z on Y it is useful to note that there are two paths from Z to Y : a directed path ($Z \rightarrow X \rightarrow Y$) and a path with a collider ($Z \rightarrow X \leftarrow C \rightarrow Y$).

³See Glynn and Quinn (2011) for a discussion in the context of registration laws and voting.

Notice that as in the path with a collider in Figure 2 (d), association is blocked along the $Z \rightarrow X \leftarrow C \rightarrow Y$ path, so that only the directed path transmits association.

The idea behind the IV strategy is that the effect of Z on Y is composed of the effect of Z on X and the effect of X on Y ,⁴ and therefore, we might be able to remove the Z on X effect, leaving the X on Y effect. However, unlike the mechanisms strategy from the previous section, the instrumental variables strategy requires additional assumptions to identify the average effect of X on Y .

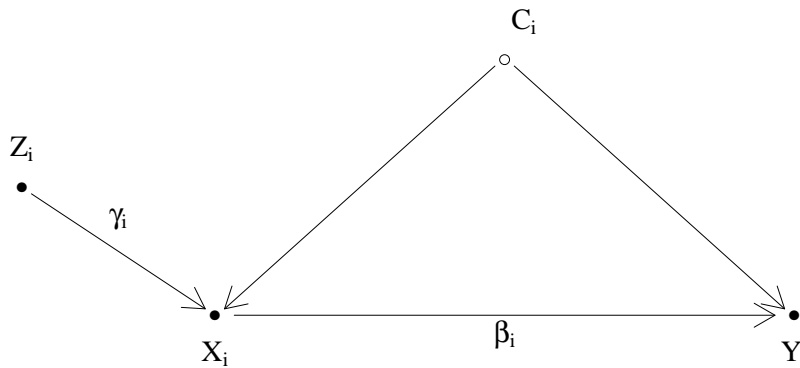


Figure 5: *DAG depicting the strategy of instrumental variables.* The variable C is the confounder, variable Z is the instrument, the variable X is treatment of interest, and the variable Y is the outcome. The parameters γ_i and β_i have been added in order to indicate heterogeneous linear effects.

The additional assumptions required for a binary instrument (Z) and a binary causal variable of interest (X) are discussed generally in Angrist et al. (1996) (see Morgan and Winship (2007) for another presentation). Here, we use a generalization of this model to a continuous X (Angrist and Pischke, 2008, pg. 187), assuming linear but heterogeneous effects. As in the previous section, we include the parameters associated with heterogeneous effects on the arrows of the graph (see Figure 5 (b)).

⁴ X is a collider on the $Z \rightarrow X \leftarrow C$ path, so this path is blocked.

To see the necessity of additional assumptions for the instrumental variables design, consider that a regression of X on Z will provide an unbiased estimate of $\mathbb{E}[\gamma_i]$, and a regression of Y on Z will provide an unbiased estimate of $\mathbb{E}[\gamma_i \cdot \beta_i]$. Recall, that the individual level effect of Z_i on Y_i on the path through X_i is just the product $\gamma_i \cdot \beta_i$, therefore, the regression of Y on Z recovers an unbiased estimate of the average of these effects. Unfortunately, the instrumental variables ratio of $\mathbb{E}[\gamma_i \cdot \beta_i]/\mathbb{E}[\gamma_i]$ will not generally recover the average effect of X on Y $\mathbb{E}[\beta_i]$, because unless γ_i and β_i are uncorrelated, $\mathbb{E}[\gamma_i \cdot \beta_i]$ does not equal $\mathbb{E}[\gamma_i] \cdot \mathbb{E}[\beta_i]$. Angrist and Pischke (2008) notes that when the effect of Z on X is monotonic (i.e., $\gamma_i \geq 0$ for all i), then the instrumental variables ratio will recover a weighted average of the effects of X on Y , with weights determined by the strength of the Z on X effect for each unit.⁵ Whether such a weighted average represents the parameter of interest will be application specific.

Returning to our exemplar, imagine that we are able to identify a factor that enhances the probability of attending the after-school ESL program, while having no direct impact on the outcome. If the after-school program occurs at the same school that students attend during the day we can suppose that some students will have an easier time attending an after-school program than others. In particular, we can anticipate that those who live far away from the school may need to take advantage of the regular bus service - which leaves at the end of the school day and does not operate a second-shift for after-school programs. If so, geographic distance from the school may serve as an instrument, predicting attendance in the after-school program. Knowing the addresses of students we may construct an instrument that satisfies the assumptions of IV analysis - assuming, e.g., that geographic distance is not also correlated with the location of highly educated parents or other potential confounders.

⁵Even if monotonicity doesn't hold, the ratio can still be interpreted in terms of an affine combination of the effects.

6 Alternate Outcomes

In the previous section, we discussed nonparametric graphical assumptions and some additional parametric assumptions that allowed the instrumental variables strategy to identify the average effect of X on Y . In this section, we demonstrate that such a combination of graphical and parametric assumptions can be used to implement another strategy based on facilitating variables that we refer to as *alternate outcomes*.⁶

The first variant of the alternate outcomes strategy utilizes an outcome that is not affected by the treatment but may be affected by the unmeasured confounder. A formalization of this procedure is presented in Figure 6 (a), with the facilitating variable (Z) as a secondary outcome that is affected by the potential confounder but not by the treatment. The key graphical assumptions are the missing arrow from X to Z and the missing arrow between Z and Y . As in the previous sections, we will additionally assume that causal effects are heterogeneous and linear, although we now extend this assumption to all of the arrows in the graph. In order to simplify the presentation of this approach, we will also assume that these effects combine additively. The structural equation model associated with this graph is presented in the Appendix A. A similar model was used in Glynn (2012) and Imai and Yamamoto (2012). Note that in this model, the average effect of X on Y can be written as $\mathbb{E}[\beta_{1i}]$.

It is straightforward to demonstrate that a regression of Y on X will identify $\mathbb{E}[\beta_{1i}] + \mathbb{E}[\beta_{2i}/\gamma_{1i}]$, and therefore, the covariational relationship will be biased by $\mathbb{E}[\beta_{2i}/\gamma_{1i}]$ (a proof is also provided in the Appendix A). It is similarly straightforward to demonstrate that a regression of Z on X will identify $\mathbb{E}[\alpha_{1i}/\gamma_{1i}]$. Therefore, when $\mathbb{E}[\beta_{2i}/\gamma_{1i}] =$

⁶This design might also be referred to as a nonequivalent dependent variables design (Cook and Campbell, 1979; Marquart, 1989; McSweeney, 1978; Minton, 1975; Reynolds and West, 1987; Shadish et al., 2001; Trochim, 1985, 1989), or a placebo test (Green et al., 2009) or a device Rosenbaum (2002). Note that a "placebo test" may also refer to other settings such as (a) where an effect is known to be zero and non-zero estimates can therefore be regarded as disconfirming evidence (Abadie et al., 2010), or (b) where meaningful variation can be found in heterogeneous treatment effects (as discussed below).

$\mathbb{E}[\alpha_{1i}/\gamma_{1i}]$, we can identify the average effect of X on Y by taking the difference between the two regression estimates. Intuitively, this approach works when the confounding is equal for the two regressions, and a special case will occur when the effect of C on Z is the same as the effect of C on Y for all individuals (i.e., when $\alpha_{1i} = \beta_{2i}$ for all i). These assumptions for identification are quite strong, but they can be relaxed in a number of ways if only partial identification is required.

In our running example, an alternate outcome might be student test scores in other subjects (other than English) covered in a high school curriculum. A student who is especially motivated or well-prepared to learn English is probably also especially motivated and well-prepared to take on other subjects. Consequently, if these confounders are responsible for the improvements in English proficiency found among those in the treatment group (relative to the control group) then we should find the same differential scores in other courses of study. That is, when we compare how members of the treatment and control group perform in math, science, and social studies we should find the treatment group out-performing the control group. If, on the other hand, we find no difference between members of the treatment and control group in other subjects we might conclude that the confounder is not present: assignment is as-if random. (The foregoing analysis is stronger if one measures rates of improvement rather than levels of performance at the end of the observation period, i.e., a pre-/post-test difference rather than a simple post-test.)

A second variant of the alternate outcomes strategy enlists an outcome that is correlated with the outcome of interest while being unaffected by the potential confounder. The formal description of this design can be found in Figure 6 (b), where the effect of X on Y is confounded by C , but the effect of X on Z is unconfounded. When this structure holds, the average effect of X on Z ($\mathbb{E}[\alpha_{1i}]$) can be estimated by regressing Z on X , and if we are willing to make assumptions about the relationship between $\mathbb{E}[\alpha_{1i}]$

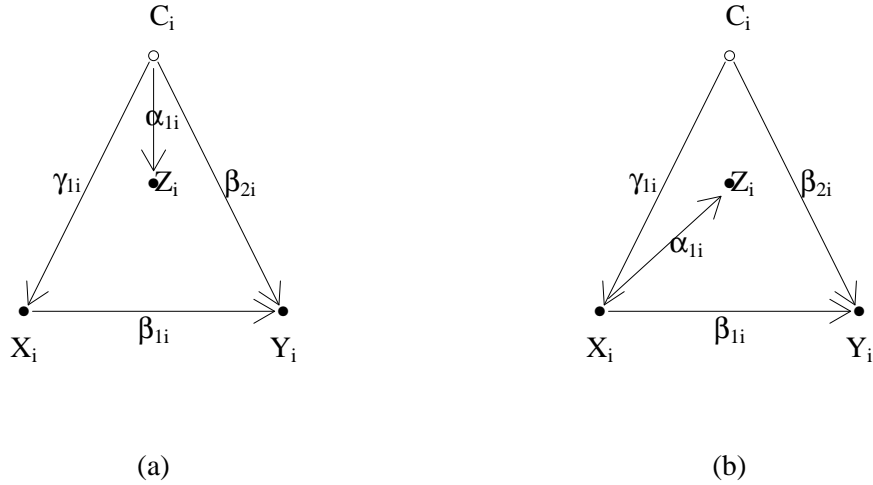


Figure 6: *DAG depicting strategies of alternate outcomes.* Parameters with i subscripts have been added in order to imply linear heterogeneous effects. Multiple effects (e.g. β_{1i} and β_{2i}) are assumed to combine additively. Panel (a) presents the strategy with a confounded outcome that is not affected by the treatment. Panel (b) presents the strategy with an unconfounded outcome that is affected by the treatment.

and $\mathbb{E}[\beta_{1i}]$, then this design will provide information about the effect of interest.

For example, we might worry that the apparent effect of an after-school ESL program on improved English proficiency is due to greater exposure to spoken English outside of class on the part of individuals that self-select into the program. This is a difficult matter to control for since we cannot easily monitor the experience of students outside of class. Suppose, however, that there is a distinctive aspect of the after-school ESL course that is not likely to be replicated in other settings: the instructor speaks with British accent. If students in the program are found to speak with this accent then we might take this as evidence that the relationship between attendance in an after-school ESL program and improvement English proficiency is not spurious, i.e., the program - rather than the potential confounder - has generated the causal effect.

In the extreme case where we are willing to assume that $\mathbb{E}[\alpha_{1i}] = \mathbb{E}[\beta_{1i}]$, then the graph in panel (b) of Figure 6 allows one to identify the average effect of X on Y .

However, it would often be more likely that we could make weaker assumptions and arrive at partial identification. For example, we may be willing to assume only that $\mathbb{E}[\alpha_{1i}] \geq \mathbb{E}[\beta_{1i}]$. In this setting, the regression of Z on X will identify $\mathbb{E}[\alpha_{1i}]$ and this will provide a lower bound on the parameter of interest $\mathbb{E}[\beta_{1i}]$.

7 Causal Heterogeneity

Causal heterogeneity – the varying impact of a causal factor, X , on units within a sample – is often regarded as noise.⁷ However, in certain situations such heterogeneity may serve as a facilitating variable. This is so in situations where the relevant moderators (Z) can be measured, in which the interaction effect of $X \cdot Z$ on Y is not subject to confounding, and in which the entire effect of X on Y is due to the interaction. In effect, the varying impact of X and Y across the sample constitutes as-if random assignment.

To formalize this approach, suppose that Z is such a moderator, and that Z has been parameterized so that when $Z = 0$, X has no effect on Y . Assume further that X and Z have a multiplicative interaction (if X and Z are both binary, then this assumption is trivially satisfied). The DAG that expresses these assumptions is presented in Figure 7. Note that we will again assume linear heterogeneous effects for the edges in the graph and additive combination rules unless otherwise specified. However, in order to accommodate the interaction, we have included the multiplicative term $Z_i \cdot X_i$ on the graph. The assumption that X has no effect when $Z = 0$ is represented by the missing edge from X_i to Y_i once the multiplicative term $Z_i \cdot X_i$ has been included in the graph. The combination of these assumptions implies that the parameter $\mathbb{E}[\beta_{3i}]$ is identified by the interaction term in the regression of Y on X , Z , and $Z \cdot X$ (a proof is provided in Appendix A). The average effect of X on Y for a particular value of

⁷Heckman (2001); Heckman and Vytlačil (2007a,b); Rhodes (2010)

$Z = z$ is $\mathbb{E}[\beta_{3i}] \cdot z$. Note that the strategy of using an interaction term from a linear regression to identify a causal effect is often used in a difference-in-difference design (see Angrist and Pischke (2008) for a textbook discussion), though in our discussion it does not require repeated measurements over time. Note as well that the assumptions stated here can be relaxed along the lines of Abadie (2005).

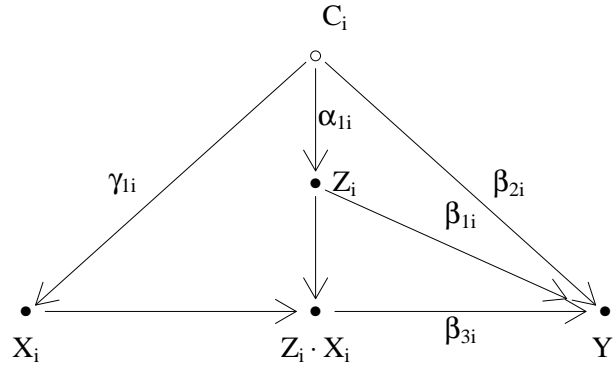


Figure 7: A DAG consistent with the strategy of causal heterogeneity. Parameters with i subscripts have been added in order to imply linear heterogeneous effects. Multiple effects (e.g. β_{1i} , β_{2i} , and β_{3i}) are assumed to combine additively although interaction terms (e.g. $Z_i \cdot X_i$) may be included explicitly in the graph.

For a concrete example we return to the hypothetical after-school ESL program, which we shall stipulate includes students with varying linguistic backgrounds and a plurality of Hispanics. Let us further stipulate that the program is geared toward the plurality group, i.e., those whose native language is Spanish. All students in the course receive the same treatment; however, by virtue of fixed characteristics (Hispanic/non-Hispanic), subjects can be expected to respond differently. In particular, one expects faster progress for Hispanic students relative to non-Hispanic students if the program is working as intended. In effect, the Hispanics in the class are treated and the non-Hispanics are not (or receive only a partial treatment). If we can assume that Hispanics

and non-Hispanics are similar in other relevant respects (i.e., in whatever ways might impact their ability to acquire English), then we can measure the average treatment effects by comparing scores among the Hispanic and non-Hispanic students who take the class. Likewise, we can narrow the comparison set to include only those with Romance-language backgrounds (Spanish versus French, Italian, Portuguese, Romanian, et al.), if we have reason to believe that the character of one's mother tongue is important in the acquisition of English.

8 Discussion

The most straightforward way to infer causation and to measure a causal effect is to observe the pattern of covariation across the two factors of theoretical interest - X and Y . This strategy is usually flawed, however, when assignment to treatment is not randomized. In nonexperimental settings there are likely to be one or more confounders, rendering X/Y covariation spurious. In this paper we have laid out five core strategies for dealing with common-cause confounders: blocking backdoors, mechanisms, instrumental variables, alternate outcomes, and causal heterogeneity. Each of these strategies enlists a third variable, Z , which plays a facilitating role in achieving identification. The distinctive role of this facilitating variable in each of these research designs is articulated through a series of causal graphs using the DAG framework developed by Judea Pearl.

In some cases our presentation formalizes strategies that are already well-known. In other cases the typology serves to highlight strategies that may be widely practiced in an informal fashion but whose methodological status remains unclear. It seems likely that some of the strategies introduced here are under-utilized in social science simply because they are not well-understood.

The presentation also serves to clarify several points of a more specific nature.

First, we have sought to show the difference in assumptions required for the different strategies. Some of these strategies— the alternate outcomes strategy and the causal heterogeneity strategy – require significantly stronger assumptions than the other two strategies. At the same time, we have demonstrated that the assumptions required for IV analysis are similar to assumptions utilized for the mechanisms strategy (when the front-door criterion does not hold).

Second, and relatedly, we recognize that these strategies are not equally promising across the varied terrain of social science. Some encounter greater difficulties (even if they can be implemented) than others. Most fall far from the experimental ideal, and thus represent second-best (or third- or fourth-best) options. The spirit of this endeavor is one of methodological pragmatism - doing the best that we can with the resources and evidence available to us at a given time (Gerring, 2011).

Third, some of the identified strategies, such as blocking back door paths and mechanisms, may be applicable to small-N samples although they require stronger assumptions or only allow certain effects to be ruled out (such as in Manski (2003)). In this respect, they apply to both quantitative and qualitative styles of analysis. In order to clarify this point let us return to our all-purpose exemplar, an afterschool ESL program. To envision the qualitative (small-N) version of the foregoing research designs, imagine that there are a small number of units in each group. Thus, in our discussion of blocking backdoor paths, the first strategy is to condition directly on the confounder, C . Suppose we have three students in the treatment group (attending the voluntary after-school ESL program) and three students in the control group (non-native speakers who attend only the regular high school ESL program). And suppose, further, that one student in each group is from a high SES background, one is from a middle SES background, and one is from a low SES background. The qualitative version of conditioning on C is to compare the two high- SES students with each other,

the two mid- SES students with each other, and the two low-SES students with each other (i.e., blocking on pre-treatment characteristics). Granted, this style of analysis is open to threats from stochastic error, and is also limited in the number of factors that can be conditioned after matching (a degrees-of-freedom problem). Even so, if certain modeling assumptions hold (e.g., the assumptions required for the paired t-test), then it may be possible to learn something about the effect of the ESL program from this analysis. Furthermore, despite the threat of stochastic error, a small-N cross-case comparison may be more informative for a subpopulation than a large-N cross-case comparison for the larger population where the population of interest is comprised of a limited number of extremely heterogeneous units, e.g., nation-states, political parties, or governmental agencies. That is, small-N approaches find their justification in precisely those settings where the large-N analog is not similar in relevant respects. Although we do not have the space to sketch out the qualitative version of each of the identified strategies, we consider this an important objective for future research. Note, however, that some strategies, such as instrumental variables have poor small sample properties (Nelson and Startz, 1990).

A final caveat concerns the limitations of causal graphs as a language of causal inference. It should be clear that, like other languages (e.g., potential outcomes), causal graphs have strengths and weaknesses. On the one hand, the visual language of causal graphs may help researchers identify new ways of approaching problems of causal inference and new sorts of evidence that would not occur to someone educated only with equations. Note that causal graphs allow one to illustrate a data generating process (DGP) - including key assumptions - that must be in place in order for an analysis to be viable. It therefore clarifies our thinking about causal inference in a specific research setting, tying together the DGP with the data model (statistical or non-statistical). On the other hand, some assumptions such as linearity or monotonicity

must be appended to the graph, while others, such as the assumption that all values of X are possible for all values of the conditioning variables, are not well represented. Finally, there are other causal quantities of potential interest, such as probabilities of causation (Tian and Pearl, 2000), pure/natural direct effects (Pearl, 2001; Robins, 2003), and principal effects (Frangakis and Rubin, 2002), that must be represented with counterfactual notation. However, causal graphs are compatible with other notational systems, such as potential outcomes, and can be used fruitfully in combination.

With these caveats noted, the typology developed in this paper places research design - rather than data analysis - front-and-center, in keeping with current trends in methodological thinking (Rubin, 2008). It should also serve to broaden the toolkit of social science. In particular, we hope to have demonstrated that disparate research designs can be understood within a unified framework through the use of causal graphs and facilitating variables.

A Mathematical Appendix

Alternate Outcomes

For the model used in Figure 6 (a) of Section 6, and for the units of analysis $i = 1, \dots, n$, we utilize Y_i to denote the outcome, X_i to denote treatment, C_i to denote the common-cause confounder, and Z_i to denote the alternate outcome. The intuition of this can be easily represented within a simple heterogeneous linear structural equations/potential-outcomes model for Y, X, Z , and C :

$$\begin{aligned}
 Y_i(x, c) &= \beta_{0i} + \beta_{1i}x + \beta_{2i}c \\
 X_i(c) &= \gamma_{0i} + \gamma_{1i}c \\
 Z_i(c) &= \alpha_{0i} + \alpha_{1i}c \\
 C_i &= \delta_i
 \end{aligned} \tag{1}$$

This model looks similar to a classical constant-effects SEM, but there are a number of important differences. First, heterogeneous effects are allowed in this model by the inclusion of the “ i ” subscripts on the coefficients. Second, error terms have been omitted because they are redundant in a model with heterogeneous intercepts (e.g., $\beta_{0i} = \beta_{0i}^* + \epsilon_i$). Third, the potential outcomes notation on the left-hand side of the equations clarifies the causal interpretation of these equations. $X_i(c)$ represents the value that X_i would take if C_i took the value c , and $Y_i(x, c)$ represents the value that Y_i would take if C_i took the value c and if X_i took the value x . Furthermore, we can use this model to define the observed values of X , Z , Y , and the potential outcome of Y in terms of only X :

$$\begin{aligned}
X_i &= \gamma_{0i} + \gamma_{1i}C_i \\
Z_i &= \alpha_{0i} + \alpha_{1i}C_i \\
Y_i &= \beta_{0i} + \beta_{1i}X_i + \beta_{2i}C_i \\
Y_i(x) &= (\beta_{0i} + \beta_{2i}C_i) + \beta_{1i}x
\end{aligned} \tag{2}$$

where the observed values of X_i and Z_i are determined by the actual (although unmeasurable) value of C_i , the observed value of Y_i is determined by the actual value of C_i and the observed value of X_i , and the potential outcome $Y_i(x)$ is the value that Y_i would take if C_i took its actual value and if X_i took the value x . Within this model, the individual-level effect of X on Y is $Y_i(x + 1) - Y_i(x) = \beta_{1i}$, and hence we might take the causal effect of interest to be the average of these individual effects, denoted as $\mathbb{E}[Y_i(x + 1) - Y_i(x)] = \mathbb{E}[\beta_{1i}]$.

Because we can write $C_i = (X_i - \gamma_{0i})/\gamma_{1i}$, the equation for Y can be rewritten in terms of only X :

$$\begin{aligned}
Y_i &= \beta_{0i} + \beta_{1i}X_i + \beta_{2i}(X_i - \gamma_{0i})/\gamma_{1i} \\
&= [\beta_{0i} - \beta_{2i}(\gamma_{0i})/\gamma_{1i}] + [\beta_{1i} + \beta_{2i}/\gamma_{1i}]X_i
\end{aligned}$$

Hence, a regression of Y on X identifies $\mathbb{E}[\beta_{1i} + \beta_{2i}/\gamma_{1i}]$ instead of the parameter of interest. Similarly, Z can be rewritten in terms of X :

$$\begin{aligned} Z_i &= \alpha_{0i} + \alpha_{1i}(X_i - \gamma_{0i})/\gamma_{1i} \\ &= [\alpha_{0i} - \alpha_{1i}(\gamma_{0i})/\gamma_{1i}] + \alpha_{1i}/\gamma_{1i}X_i \end{aligned}$$

Therefore, a regression of Z on X identifies $\mathbb{E}[\alpha_{1i}/\gamma_{1i}]$. Depending on the relationship between α_{1i} and β_{2i} , it may be possible to use a regression of Z on X to correct the bias in the regression of Y on X . For example, if $\mathbb{E}[\beta_{2i}/\gamma_{1i}] = \mathbb{E}[\alpha_{1i}/\gamma_{1i}]$, then the two regression estimates can be differenced in order to identify $\mathbb{E}[\beta_{1i}]$.

Causal Heterogeneity

The model presented in Figure 7 in Section 7 can be represented as the following structural equation model:

$$\begin{aligned} Y_i(x, z, c) &= \beta_{0i} + \beta_{1i}z + \beta_{2i}c + \beta_{3i} [x \cdot z] \\ [X_i \cdot Z_i] (x, z) &= x \cdot z \\ X_i(c) &= \gamma_{0i} + \gamma_{1i}c \\ Z_i(c) &= \alpha_{0i} + \alpha_{1i}c \\ C_i &= \delta_{0i} \end{aligned}$$

Note that for this model, the absence of a direct effect of X on Y implies that a unit change in X results in an effect of $\beta_{3i} \cdot Z_i$ change in Y . Therefore under the modeling assumption that β_{3i} and Z_i don't covary, the average effect of a unit change in X on Y is $\mathbb{E}[\beta_{3i}] \cdot \mathbb{E}[Z_i]$.

The observed values of Y can be written as the following:

$$Y_i = \beta_{0i} + \beta_{1i}Z_i + \beta_{2i}C_i + \beta_{3i} [X_i \cdot Z_i]$$

So as long as the following four conditions hold,

$$\mathbb{E}[\beta_{0i}|X_i = x, Z_i = z] = \mathbb{E}[\beta_{0i}]$$

$$\mathbb{E}[\beta_{1i}Z_i|X_i = x, Z_i = z] = \mathbb{E}[\beta_{1i}] \cdot z$$

$$\mathbb{E}[\beta_{2i}C_i|X_i = x, Z_i = z] = \nu + \xi x + \zeta z$$

$$\mathbb{E}[\beta_{3i}[X_i \cdot Z_i]|X_i = x, Z_i = z] = \mathbb{E}[\beta_{3i}] \cdot x \cdot z$$

the parameter $\mathbb{E}[\beta_{3i}]$ can be identified with the coefficient on the interaction term for the following regression,

$$\mathbb{E}[Y_i|X_i = x, Z_i = z] = \mathbb{E}[\beta_{0i}] + \mathbb{E}[\beta_{1i}]z + \nu + \xi x + \zeta z + \mathbb{E}[\beta_{3i}]x \cdot z,$$

and the average effect can be identified with $\mathbb{E}[\beta_{3i}] \cdot \mathbb{E}[Z_i]$. Notice that these conditions allow for the possibility that both X and Z are confounded. However, it is critical that X not have a direct effect on Y and that there be no interactive relationship between X and Z in the regression for the unmeasured confounder.

References

- Abadie, Alberto. 2005. “Semiparametric Difference-in-Difference Estimators.” *Review of Economic Studies* 72:1 – 19.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program.” *Journal of the American Statistical Association* 105:493 – 505.
- Abbott, Andrew. 1996a. “Conceptions of Time and Events in Social Science Methods: Causal and Narrative Approaches.” *Historical Methods* 23(4):140 – 150.
- Abbott, Andrew. 1996b. “From Causes to Events: Notes on Narrative Positivism.” *Sociological Methods and Research* 20(4):428 – 455.
- Abrams, Philip. 1982. *Historical Sociology*. Ithaca: Cornell University Press.
- Aminzade, Ronald. 1992. “Historical Sociology and Time.” *Sociological Methods and Research* 20(4):456–480.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association* 91(434):444 – 455.

- Angrist, Joshua D., and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricists Companion*. Princeton University Press.
- Barton, Alan H., and Paul F. Lazarsfeld. 1969. *Some Functions of Qualitative Analysis in Social Research*, Reading, MA: Addison-Wesley.
- Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry Weingast. 1998. *Analytic Narratives*. Princeton: Princeton University Press.
- Bennett, Andrew. 1999. "Causal Inference in Case Studies: From Mills Methods to Causal Mechanisms." In *Paper presented at the annual meeting of the American Political Science Association*. Atlanta, GA.
- Bhaskar, Roy. 1975/1978. *A Realist Theory of Science*. Sussex: Harvester Press.
- Brown, Robert. 1984. *The Nature of Social Laws: Machiavelli to Mill*. Cambridge: Cambridge University Press.
- Campbell, Donald T. 1975. "Degrees of Freedom and the Case Study." *Comparative Political Studies* 8(8):178–193.
- Collier, David, and James Mahoney. 1996. "Insights and Pitfalls: Selection Bias in Qualitative Research." *World Politics* 49(1):56–91.
- Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Wadsworth Publishing.
- Elwert, Felix, and Christopher Winship. 2012. "Endogenous Selection Bias." Unpublished Manuscript.
- Frangakis, Constantin E., and Donald B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58(1):21–29.
- George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences (Belfer Center Studies in International Security)*. The MIT Press.
- George, Alexander L., and Timothy J. McKeown. 1985. *Case Studies and Theories of Organizational Decision Making*. Santa Barbara: JAI Press.
- Gerring, John. 2011. "How Good is Good Enough? A Multidimensional, Best-Possible Standard for Research Design." *Political Research Quarterly* 64(3):625–36.
- Glynn, Adam. 2012. "The Product and Difference Fallacies for Indirect Effects." *American Journal of Political Science* 56(1):257–269.
- Glynn, Adam, and Kevin Quinn. 2011. "Why Process Matters for Causal Inference." *Political Analysis* 19(3):273–286.
- Goldstone, Jack A. 1997. "Methodological Issues in Comparative Macrosociology." *Comparative Social Research* 16:121–132.
- Green, D P, T Y Leong, H L Kern, A S Gerber, and C W Larimer. 2009. "Testing the accuracy of regression discontinuity analysis using experimental benchmarks." *Political Analysis* 17(4):400.

- Griffin, Larry J. 1992. "Temporality, Events, and Explanation in Historical Sociology: An Introduction." *Sociological Methods and Research* 20(4):403–427.
- Griffin, Larry J. 1993. "Narrative, Event-Structure Analysis, and Causal Interpretation in Historical Sociology." *American Journal of Sociology* 98:1094–1133.
- Harre, Rom. 1972. *The Philosophies of Science*. London: Oxford University Press.
- Heckman, James J. 2001. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: The Nobel Lecture." *Journal of Political Economy* 109:673–748.
- Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66:1017–1098.
- Heckman, James J., and Edward Vytlacil. 2007a. *Econometric Evaluation of Social Programs. Part I: Causal Models, Structural Models and Econometric Policy Evaluations*, Amsterdam: North Holland Press.
- Heckman, James J., and Edward Vytlacil. 2007b. *Econometric Evaluation of Social Programs. Part II: Using the Marginal Treatment Effect to Evaluate Social Programs, and to Forecast their Effects in New Environments*, Amsterdam: North Holland Press.
- Hernan, Miguel, and James Robins. 2012. *Causal Inference*. Chapman & Hall/CRC.
- Imai, Kosuke, and Teppei Yamamoto. 2012. "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments." Unpublished manuscript.
- Imbens, Guido W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *The Review of Economics and Statistics* 86(1):4–29.
- Katznelson, Ira. 1997. *Structure and Configuration in Comparative Politics*, Cambridge: Cambridge University Press, pp. 81–112.
- Kiser, Edgar. 1996. "The Revival of Narrative in Historical Sociology: What Rational Choice Can Contribute." *Politics and Society* 24:249–271.
- Knight, Carly, and Christopher Winship. 2012. "The Causal Implications of Mechanistic Thinking." Unpublished Manuscript.
- Komarovsky, Mirra. 1940. *The Unemployed Man and His Family: The Effect of Unemployment upon the Status of the Man in Fifty-nine Families*. New York: Dryden Press.
- Lalonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review* 76(4):604–620.
- Little, Daniel. 1998. *Microfoundations, Method, and Causation*. New Brunswick: Transaction.
- Manski, Charles F. 2003. *Partial identification of probability distributions*. Springer.
- Marquart, J. 1989. "A pattern matching approach to assess the construct validity of an evaluation instrument." *Evaluation and Program Planning* 12(1):37–43.

- McMullin, Ernan. 1984. *Two Ideals of Explanation in Natural Science*, Minneapolis: University of Minnesota Press.
- McSweeney, A.J. 1978. "The Effects of Response Cost on the Behavior of a Million Persons: Charging for Directory Assistance in Cincinnati." (11):47–51.
- Mill, J. S. 1843/1872. *A System of Logic, 8 ed.*. London: Longmans, Green.
- Mink, Louis. 1987. *History and Fiction as Modes of Comprehension*, Ithaca: Cornell University Press.
- Minton, J.H. 1975. "The Impact of Sesame Street on Reading Readiness of Kindergarten Children." *Sociology of Education* 48:141–151.
- Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York, NY: Cambridge University Press.
- Nelson, Charles R., and Richard Startz. 1990. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica* 58(4):967–976.
- Pearl, Judea. 2000. *Causality: Models, Reasoning and Inference*. Cambridge University Press.
- Pearl, Judea. 2001. "Direct and Indirect Effects." In Proceedings of the American Statistical Association Joint Statistical Meetings, Minn. MN.
- Pearl, Judea. 2009. *Causality: Models, Reasoning and Inference*. Cambridge University Press, second edition.
- Quadagno, Jill, and Stan J. Knapp. 1992. "Have Historical Sociologists Forsaken Theory?: Thoughts on the History/Theory Relationship." *Sociological Methods and Research* 20:481–507.
- Reynolds, K.D., and S.G. West. 1987. "A Multiplist Strategy for Strengthening Nonequivalent Control Group Designs." *Evaluation Review* 11:691–714.
- Rhodes, William. 2010. "Heterogeneous Treatment Effects: What Does a Regression Estimate?" *Evaluation Review* 34(4):334–361.
- Roberts, Clayton. 1996. *The Logic of Historical Explanation*. University Park: Pennsylvania State University Press.
- Robins, James M. 2003. *Semantics of Causal DAG Models and The Identification of Direct and Indirect Effects*, Oxford University Press.
- Rosenbaum, Paul R. 2002. *Observational Studies (Springer Series in Statistics)*. New York: Springer.
- Rosenbaum, Paul R. 2009. *Design of Observational Studies (Springer Series in Statistics)*. New York: Springer.
- Roth, Paul A. 1994. *Narrative Explanations: The Case of History*, Cambridge: MIT Press.

- Rubin, Donald B. 2008. "For Objective Causal Inference, Design Trumps Analysis." *Annals of Applied Statistics* 2(3):808–840.
- Rueschemeyer, Dietrich, and John D. Stephens. 1997. "Comparing Historical Sequences A Powerful Tool for Causal Analysis." *Comparative Social Research* 16:55–72.
- Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Sewell, Jr., William H. 1992. "Introduction: Narratives and Social Identities." *Social Science History* 16:479–488.
- Sewell, Jr., William H. 1996. *Three Temporalities: Toward an Eventful Sociology*, Ann Arbor: University of Michigan Press, pp. 245–280.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Publishing.
- Somers, Margaret R. 1992. "Narrativity, Narrative Identity, and Social Action: Rethinking English Working-Class Formation." *Social Science History* 16:591–630.
- Stone, Lawrence. 1979. "The Revival of Narrative: Reflections on a New Old History." *Past and Present* 85:3–24.
- Stryker, Robin. 1996. "Beyond History versus Theory: Strategic Narrative and Sociological Explanation." *Sociological Methods and Research* 24(3):304–352.
- Tian, Jin, and Judea Pearl. 2000. "Probabilities of Causation: Bounds and Identification." *Annals of Mathematics and Artificial Intelligence* 28:287–313.
- Tian, Jin, and Judea Pearl. 2002. "A General Identification Condition for Causal Effects." In *Proceedings of the American Association of Artificial Intelligence*.
- Trochim, W. M.K. 1985. "Pattern Matching, Validity, and Conceptualization in Program Evaluation." *Evaluation Review* 9(5):575–604.
- Trochim, William M K. 1989. "Outcome pattern matching and program theory." *Evaluation and Program Planning* 12(4):355–366.
- Vanderweele, Tyler J. 2009. "On the Relative Nature of Overadjustment and Unnecessary Adjustment." *Epidemiology* 20(4):496–499.
- Watkins, J.W.N. 1994. *Historical Explanation in the Social Sciences*, Cambridge: MIT Press, pp. 441–450.
- Winship, Christopher, and David J. Harding. 2008. "A Mechanism-Based Approach to the Identification of Age-Period-Cohort Models." *Sociological Methods & Research* 36(3):362–401.