Chapter 13
Graphical Causal Models

Felix Elwert

Abstract This chapter discusses the use of directed acyclic graphs (DAGs) for causal inference in the observational social sciences. It focuses on DAGs' main uses, discusses central principles, and gives applied examples. DAGs are visual representations of qualitative causal assumptions: They encode researchers' beliefs about how the world works. Straightforward rules map these causal assumptions onto the associations and independencies in observable data. The two primary uses of DAGs are (1) determining the identifiability of causal effects from observed data and (2) deriving the testable implications of a causal model. Concepts covered in this chapter include identification, d-separation, confounding, endogenous selection, and overcontrol. Illustrative applications then demonstrate that conditioning on variables at any stage in a causal process can induce as well as remove bias, that confounding is a fundamentally causal rather than an associational concept, that conventional approaches to causal mediation analysis are often biased, and that causal inference in social networks inherently faces endogenous selection bias. The chapter discusses several graphical criteria for the identification of causal effects of single, time-point treatments (including the famous backdoor criterion), as well identification criteria for multiple, time-varying treatments.

Introduction

Visual representations of causal models have a long history in the social sciences, first gaining prominence with path diagrams for linear structural equation models in the 1960s (Blalock 1964; Duncan 1975). Since these beginnings, methodologists in various disciplines have made remarkable progress in developing formal theories for graphical causal models that not only generalize the linear path diagrams of yore into a fully nonparametric framework but also integrate graphical models with the reigning potential outcomes framework of causal inference. Best of all, methodologists have developed a system that is both rigorous and easy to use.

In recent years, graphical causal models have become largely synonymous with *directed acyclic graphs* (DAGs). On their own, DAGs are just mathematical objects built from dots and arrows. With a few assumptions, however, DAGs can be rigorously related both to data (probability distributions) and to causal frameworks, including the potential outcomes framework. Various closely related (but not identical) bridges between DAGs and causation exist (see Robins and Richardson (2011) for a concise

F. Elwert (✉)
Department of Sociology, Center for Demography and Ecology, University of Wisconsin–Madison, Madison, WI, USA
e-mail: elwert@wisc.edu

S.L. Morgan (ed.), *Handbook of Causal Analysis for Social Research*,
Handbooks of Sociology and Social Research, DOI 10.1007/978-94-007-6094-3_13,
© Springer Science+Business Media Dordrecht 2013

comparison). Among these, the interpretation of DAGs as nonparametric structural equation models (NPSEM) unquestionably dominates the literature. This chapter discusses the use of DAGs interpreted as NPSEM for causal inference (henceforth simply called DAGs) in the observational social sciences. It focuses on DAGs' main uses, building powerful rules from basic principles, and it gives applied examples. Technical details are found in the specialist literature.

DAGs are visual representations of qualitative causal assumptions: They encode researchers' expert knowledge and beliefs about how the world works. Simple rules then map these causal assumptions onto statements about probability distributions: They reveal the structure of associations and independencies that could be observed if the data were generated according to the causal assumptions encoded in the DAG. This translation between causal assumptions and observable associations underlies the two primary uses for DAGs. First, DAGs can be used to prove or disprove the identification of causal effects, that is, the possibility of computing causal effects from observable data. Since identification is always conditional on the validity of the assumed causal model, it is fortunate that the second main use of DAGs is to present those assumptions explicitly and reveal their testable implications, if any.

DAGs are rigorous tools with formal rules for deriving mathematical proofs. And yet, in many situations, using DAGs in practice requires only modest formal training and some elementary probability theory. DAGs are thus extremely effective for presenting hard-won lessons of modern methodological research in a language comprehensible to applied researchers. Beyond this pedagogical use, DAGs have become an enormously productive engine of methodological progress in their own right. The rapid adoption of DAGs across disciplines in recent years, especially in epidemiology, testifies to their success.

DAGs were primarily developed in computer science by Judea Pearl (1985, 1988, 1995, [2000] 2009) and Spirtes et al. ([1993] 2001), with important contributions by statisticians, philosophers, mathematicians, and others, including Verma, Lauritzen, Balke, Tian, Robins, Greenland, Hernán, Shpitser, and VanderWeele. For a detailed technical treatment, see Pearl (2009) and the references therein. For recent, less-technical overviews, see Morgan and Winship (2007), Pearl (2010, 2012a), and the excellent chapter by Glymour and Greenland (2008). For important early applications in epidemiology, see Greenland et al. (1999a), Robins (2001), Cole and Hernán (2002), and Hernán et al. (2004). For sociological applications, see Morgan and Winship (2007, 2012), Elwert and Winship (forthcoming), Winship and Harding (2008), Shalizi and Thomas (2011), Sharkey and Elwert (2011), and Wodtke et al. (2011).

This chapter has two overarching aims: first, to establish fundamental concepts and rules for using DAGs and second, to provide applied social science examples, conceptual insights, and extensions. In the first half of the chapter, I begin by emphasizing the difference between identification and estimation. I then introduce basic graphical terminology and the three structural sources of observable associations (as well as three corresponding biases). A section on d-separation consolidates the three sources of association into a single tool for translating between causation and association and illustrates how to derive the testable implications of a causal model. Following a short interlude on NPSEM and effect heterogeneity, I present seven interrelated graphical identification criteria, including the adjustment criterion and the famous backdoor criterion. In the second half of the chapter, I demonstrate that confounding is a causal concept that cannot be reduced to associational rules, use DAGs to elucidate diverse examples of selection bias at all stages of the causal process, illustrate the central problem of causal mediation analysis, and show how DAGs can illuminate causal inference in social network analysis. The final section illustrates a powerful graphical identification criterion for the causal effects of time-varying treatments.

Identification and Estimation

Causal inference must bridge a gap between goals and means. Analysts seek causation, but the data, on their own, only communicate associations. Associations usually consist of a mixture of causal and noncausal (spurious) components. *Identification analysis* determines whether, and under which conditions, it is possible to strip an observed association of all its spurious components. We say that a causal effect is *identified* if a properly stripped association equals ("identifies") the causal effect.

Identification analysis requires causal assumptions about how the data were generated. The sum of these causal assumptions is called a *causal model*, which must describe both how the world works (how observed and unobserved variables take their values) and how the data were collected (what variables and variable values are recorded) (Greenland 2010). All identification results are conditional on the validity of the stated causal model (Pearl 1995).[1] DAGs are useful for identification analysis because they are transparent graphical displays of the causal model. The relative ease with which causal assumptions and their implications can be encoded in, and read from, a DAG enables subject-matter experts to assess and debate their validity (while acknowledging that it is never possible to test all relevant causal assumptions in nonexperimental studies (Robins and Wasserman 1999)).

Identification is not the same as estimation. Identification refers to the *possibility* of correctly estimating a causal effect asymptotically from a given set of observed variables, as the number of observations goes to infinity. Actually estimating, that is, computing, the causal effect from finite sample data is a different matter. As we will see, causal effects that are nonparametrically identified (i.e., asymptotically computable regardless of the distribution of the variables and the functional form of the causal effects) often cannot be estimated with specific conventional parametric regression models.

Relatedly, it is important to distinguish between causal models (e.g., as encoded in DAGs) and statistical models (e.g., regression equations). In theory, nonparametric identification immediately implies a valid general nonparametric estimator. In practice, however, the nonparametric estimator is often not feasible with real data such that analysts must try their luck with more restrictive parametric estimators. It is often far from obvious how one might specify a parametric statistical model to estimate the parameters of a causal model. Conversely, it is impossible to offer a causal interpretation of a statistical model absent an explicitly stated causal model: The same regression coefficient may yield drastically different interpretations depending on which causal model the analyst believes to be true. The common practice of writing a causal model in regression-like algebraic notation, or, worse, of writing regression equations in lieu of specifying an explicit causal model, can lead to serious confusion. DAGs are helpful for maintaining the distinction between causal and statistical models because they offer a notation specifically for causal assumptions.

The graphical approach to causal inference reviewed in this chapter appeals to the same counterfactual notion of causality as the potential outcomes framework of causality (e.g., Neyman [1923] 1990; Rubin 1974; Holland 1986). Analysts should view graphical and potential outcome notations as equivalent: A theorem in one framework is a theorem in the other (Galles and Pearl 1998). The choice between the two notations is to an extent a matter of taste. For many purposes (such as articulating assumptions and inferring testable implications), DAGs are more accessible than algebraic notation.

[1] Identification is also relative to the set of observed variables. Identification may be possible for one set of observed variables, but not for another set. Mimicking the logic of secondary data analysis, here I assume that the analyst is given a set of observed variables (and hence that all other variables are unobserved). Identification analysis can also be used to ask what sets of variables *should* be observed to achieve identification.

For other purposes, a hybrid potential outcomes/graphical notation is more helpful. Of course, clarity is in the eye of the beholder. Pearl (2009: Chapter 7, 2012a) discusses equivalence and trade-offs.[2]

Terminology and Preliminaries

DAGs consist of three elements: variables (nodes, vertices), arrows (edges), and missing arrows. *Arrows* represent possible *direct causal effects* between pairs of variables and order the variables in time. The arrow between C and Y in Fig. 13.1 means that C may exert a direct causal effect on Y for at least one member of the population. *Missing arrows* represent the strong assumption of no direct causal effect between two variables for every member of the population (a so-called "strong null" hypothesis of no effect). The missing arrow between T and Y asserts the complete absence of a direct causal effect of T on Y. DAGs are nonparametric constructs: They make no statement about the distribution of the variables (e.g., normal, Poisson), the functional form of the direct effects (e.g., linear, nonlinear, stepwise), or the magnitude of the causal effects.

The variables directly caused by a given variable are called its *children*. The only child of T is C. All variables directly or indirectly caused by a given variable are called its *descendants*. The descendants of T are C and Y. The direct causes of a variable are called its *parents*. The only parent of T is X. All direct and indirect causes of a variable are called its *ancestors*. The ancestors of T are X, U_1, and U_2.

Paths are sequences of adjacent arrows that traverse any given variable at most once. The arrows along a path may point in any direction. *Causal paths* are paths in which all arrows point away from the treatment and toward the outcome; all other paths are called *noncausal paths*. Causal and noncausal paths are defined relative to a specific treatment and outcome. If T is the treatment and Y is the outcome, then among the eight distinct paths between T and Y, $T \rightarrow C \rightarrow Y$ is the only causal path, and $T \rightarrow C \leftarrow X \leftarrow U_1 \rightarrow Y$ is a noncausal path.

A *collider* on a path is a variable with two arrows along the path pointing into it. Otherwise, the variable is a *noncollider* on the path. Note that the same variable may be a collider on one path and a noncollider along another. For example, X is a collider along the path $U_1 \rightarrow X \leftarrow U_2$ and a noncollider along the path $U_1 \rightarrow X \rightarrow T$.

DAGs encode the analyst's qualitative causal assumptions about the data generating process in the population. But in contrast to conventional social theory, which focuses on justifying what relationships do exist, DAGs insistently redirect the analyst's attention to justifying what arrows do not exist. Present arrows represent the analyst's ignorance. Missing arrows, by contrast, represent definitive claims of knowledge. It is the *missing arrows*—"exclusion restrictions" in the language of economics—that enable the identification of a causal effect. Adding arrows to an existing set of variables in a DAG (i.e., relaxing exclusion restrictions) never aids nonparametric identification.

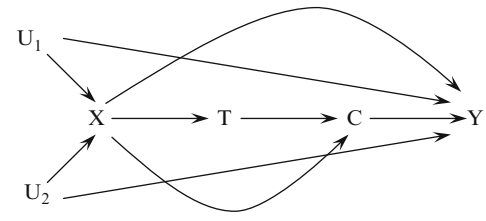

Fig. 13.1 A directed acyclic graph (*DAG*)

[2] A detailed tutorial for reading counterfactuals (including nested counterfactuals) off a DAG is presented in Section 4.4 of Pearl (2012a).

Adding variables, however, can help (Pearl 2009). Being representations of the population, DAGs abstract from sampling variability and thus from many issues in statistical inference, such as standard errors and significance tests.

When working with DAGs, the analyst (for the most part) needs to assume that the DAG captures the causal structure of everything that matters about a process. What matters most are the common causes. *Causal DAGs* are defined as DAGs that include all common measured and unmeasured causes of any pair of variables already included in the DAG (Spirtes et al. [1993] 2001). All DAGs in this chapter are assumed to be causal DAGs. My assertion that the DAG in Fig. 13.1 is a causal DAG means that I believe, for example, that there exists no variable Z that exerts both a direct causal effect on X and a direct causal effect on T. Causal DAGs may include variables that are not common causes, such as T. As a conventional shorthand, causal DAGs usually do not display the idiosyncratic factors generically assumed to affect each variable in the DAG, that is, exogenous non-common causes (sometimes called *independent error terms*),[3] such as $U_T \rightarrow T$. Such idiosyncratic causes never contribute to the ability to nonparametrically identify causal effects. In some cases, it is useful to display idiosyncratic error terms and other variables that are not common causes in order to demonstrate why identification may *fail*, and I will do so on occasion in the examples below. Asserting that a DAG represents a causal DAG is a bold claim—a claim that should be carefully considered and tested against the data as far as possible. As we will see below, a causal model encoded in a causal DAG straightforwardly enables the enumeration of its testable components.

DAGs are called "acyclic" because they may not contain directed cycles, that is, paths that can be traced strictly along the direction of the arrows to arrive back at the starting point. Acyclicity preserves the truth that the future cannot cause the past. Apparent counterexamples are usually resolved by more finely articulating the temporal sequence of events (Greenland et al. 1999a). For example, the statement that "schooling and earnings cause each other" might be understood to mean that, say, a 15-year-old student's expectations about her future earnings may influence her decision to enter college at age 18, and graduating from college at age 21, in turn, may increase her wages at age 30. Youthful wage expectations are not the same thing as adult earnings, mandating that they be represented as separate variables in the DAG, which removes the apparent cycle. Most theoretical and practical applications of DAGs in the literature assume that true simultaneity (of A causing B and B causing A) does not exist, but theory exists for cyclic graphs as well.

The Three Sources of Association: Causation, Confounding, and Endogenous Selection

With the help of two mild assumptions, analysts can translate from the causal assumptions encoded in the DAG to associations observable in the data.[4] The rules for moving from causation to association are remarkably straightforward. Absent chance (i.e., apart from sampling variation), one only needs to consider the three elementary causal structures from which all DAGs can be constructed: chains $A \rightarrow C \rightarrow B$ (and its contraction $A \rightarrow B$), forks $A \leftarrow C \rightarrow B$, and inverted forks $A \rightarrow C \leftarrow B$. Conveniently, these structures correspond exactly to causation, confounding, and endogenous selection.

[3] By contrast, marginally correlated error terms must be explicitly included in the causal DAG, since they represent common causes.

[4] These assumptions are, first, the *causal Markov assumption*, which states that a variable is independent of its nondescendants given its parents, and second, *stability* or *faithfulness*, which, among other things, rules out exact cancelation of positive and negative effects. In this chapter, I mostly use *weak faithfulness*, which is the reason for interpreting arrows as *possible* rather than *certain* direct effects. Glymour and Greenland (2008) give an accessible summary. See Pearl (2009) and Spirtes et al. ([1993] 2001) for technical details.

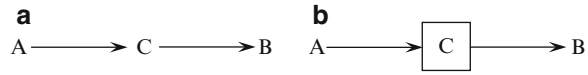

Fig. 13.2 (**a**) *A* and *B* are associated by causation. This marginal association identifies the causal effect of *A* on *B*. (**b**) *A* and *B* are conditionally independent given *C*. This conditional association does not identify the causal effect of *A* on *B* (overcontrol bias)

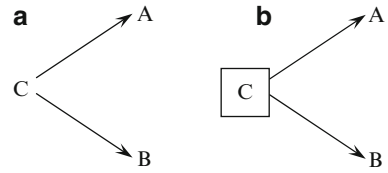

Fig. 13.3 (**a**) *A* and *B* are associated by common cause. The marginal association does not identify the causal effect of *A* on *B* (confounding bias). (**b**) *A* and *B* are conditionally independent given *C*. The conditional association does identify the causal effect of *A* on *B*

First, two variables may be marginally associated if one variable directly or indirectly *causes* the other. In Fig. 13.2a, *A* and *B* are associated only because *A* is an indirect cause of *B*. The observed marginal association between *A* and *B* identifies the causal effect of *A* on *B*. Conditioning on *C*—for example, by including *C* as a control variable in a nonparametric regression of *A* on *B* and *C*—would block, or control away, the association flowing from *A* to *B* (Fig. 13.2b). Thus, the conditional association between *A* and *B* given *C* would not identify the causal effect of *A* on *B*. We say that conditioning on *C* leads to *overcontrol bias* (Elwert and Winship forthcoming). A box drawn around a variable denotes conditioning.

Second, two variables may be associated if they share a common cause. For example, *A* and *B* in Fig. 13.3a are only associated because they are both caused by *C*. This is the familiar situation of *common cause confounding bias* (*confounding* for short). The marginal association between *A* and *B* is spurious, or biased, because it does not identify a causal effect of *A* on *B*. Conditioning on *C* would eliminate this spurious association (Fig. 13.3b). Therefore, the conditional association between *A* and *B* given *C* would identify the causal effect of *A* on *B* (which, in this DAG, happens to be zero).[5]

The third way in which two variables may be associated is less well known, but it is no less important: Conditioning on the common outcome of two variables (i.e., a collider) induces a spurious association between them for at least one value of the collider. *A* and *B* in Fig. 13.4a are marginally independent because they do not cause each other and do not share a common cause. Thus, the marginal association between *A* and *B* identifies the causal effect (the marginal association and the causal effect are both zero). Conditioning on the common outcome *C* of *A* and *B*, however, induces a nonzero association between *A* and *B* (Fig. 13.4b). The conditional association between *A* and *B* given *C* does not identify the causal effect of *A* on *B*. Elwert and Winship (forthcoming) call this phenomenon *endogenous selection*.[6] A dashed line between two variables, *A*- - -*B* (without arrowheads), indicates an association induced by endogenous selection. Dashed lines act like a regular path segment.

[5]Common cause confounding by unobserved variables is sometimes represented by a bi-headed dashed arrow.

[6]Terminology is in flux. The name "endogenous selection bias" highlights that the problem originates from conditioning on an endogenous variable. Others prefer "selection bias" (Hernán et al. 2004), "collider stratification bias" (Greenland 2003), "M-bias" (Greenland 2003), "Berkson's [1946] bias," "explaining away effect" (Kim and Pearl 1983), or "conditioning bias" (Morgan and Winship 2007). Simpson's paradox (Hernán et al. 2011) and the Monty-Hall dilemma (Burns and Wieth 2004) involve specific examples of endogenous selection. The shared structure of some examples of

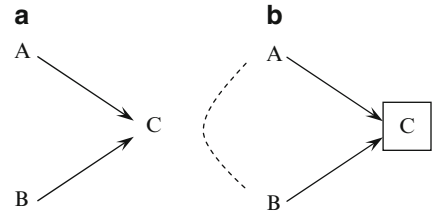

Fig. 13.4 (**a**) *A* and *B* are marginally independent. The marginal association identifies the causal effect of *A* on *B*. (**b**) *A* and *B* are associated due to conditioning on a common outcome (collider). The conditional association between *A* and *B* given *C* does not identify the causal effect of *A* on *B* (endogenous selection bias)

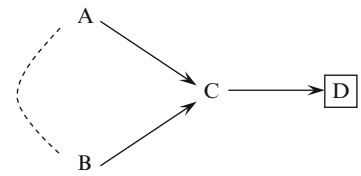

Fig. 13.5 Conditioning on a descendant, *D*, of a collider also induces a spurious association between *A* and *B*

Conditioning on the descendant of a collider results in the same problem as conditioning on the collider itself. For example, in Fig. 13.5, conditioning on *D* (rather than *C*) may induce a spurious association between *A* and *B*. The intuition for this fact is that *D* carries information about *A* and *B* as encoded in *C*, such that conditioning on *D* amounts to conditioning on *C* itself.

Endogenous selection is a common problem in the social sciences—a problem virtually guaranteed to occur if the analyst conditions on a collider variable.[7] To develop intuition for the problem, consider the following causal model for the relationships between productivity, *A*, originality, *B*, and academic tenure, *C*. Suppose, for argument's sake, that productivity and originality are unassociated in the general population (i.e., productivity does not cause originality, originality does not cause productivity, and productivity and originality do not share any common cause). Suppose further that originality and productivity are separately sufficient for promotion to tenure. The causal DAG for this model is given by $A \rightarrow C \leftarrow B$. Tenure is a collider variable. Now condition on the collider, for example, assess the relationship between originality and productivity only among tenured faculty. Knowing that an unoriginal scholar has tenure implies that he must have been productive. Conversely, knowing that an unproductive scholar has tenure implies that he must have been original. Either way, conditioning on the collider tenure creates an association between productivity and originality among tenured faculty, even though one does not cause the other.

This example may lack sociological nuance, but it demonstrates the essential logic of endogenous selection bias. To achieve greater realism, one could embellish the causal model by loosening assumptions to allow that *B* causes *A* (maybe because original scholars have more to write about) and that *A* and *B* share a common cause (maybe because having had a good graduate school advisor has caused both a knack for original thinking and irreproachable work habits). The problem of endogenous selection, however, would not go away. The observed conditional association between productivity

endogenous selection bias has been known in the social sciences at least since Heckman (1976). For a comprehensive treatment, see Elwert and Winship (forthcoming).

[7]Endogenous selection bias is guaranteed if one assumes that positive and negative arrows do not cancel each other out exactly, i.e. if the DAG is faithful. Faithfulness is a mild assumption since exact cancelation is exceedingly unlikely in practice.

and originality given tenure would remain biased for the true causal effect of productivity on originality—it would represent a mixture of (1) the true causal effect of originality on productivity, (2) confounding by advisor quality, and (3) the spurious association between productivity and originality induced by conditioning on tenure.

In sum, there are three structural sources of association and three corresponding structural sources of bias. It is helpful to draw sharp distinctions between these biases because they originate from different causal structures and from different analytic actions and hence require different remedies. Confounding bias arises from failure to condition on a common cause; the remedy is to condition on the common cause. Overcontrol bias results from conditioning on a variable on a causal path between treatment and outcome; the remedy is not to condition on the variable. Endogenous selection bias results from conditioning on a (descendant of a) collider on any path connecting treatment and outcome; the remedy is not to condition on such variables.[8]

d-Separation and Testable Implications

All associations between variables in a DAG are transmitted along paths. Not all paths, however, transmit association. Whether a path transmits association depends both on the orientation of its arrows and on which variables the analyst conditions on. The concept of *d-separation* consolidates the three sources of association—causation, confounding, and endogenous selection—into a general graphical rule to determine when a path transmits association and when it does not.

D-separation (Pearl 1988): A path between two variables, A and B, is said to be *d-separated* (blocked or closed) if:

1. The path contains a noncollider that has been conditioned on, for example, $A \rightarrow \boxed{C} \rightarrow B$ or $A \leftarrow \boxed{C} \rightarrow B$; or
2. The path contains a collider that has not been conditioned on, for example, $A \rightarrow C \leftarrow B$, and no descendant of any collider on the path has been conditioned on either.

A path is said to be *d-connected (unblocked or open)* if it is not d-separated. We say that two (sets of) variables are d-separated if they are d-separated along all paths; they are d-connected otherwise. In an important theorem, Verma and Pearl (1988; Pearl 2009: 18) prove that if two (sets of) variables A and B are d-separated by conditioning on a (possibly empty) set of variables C in a causal DAG, then A is statistically independent of B conditional on C, $A \perp\!\!\!\perp B|C$ (where $\perp\!\!\!\perp$ stands for statistical independence) in any distribution generated by a process consistent with the DAG (so-called compatible distributions). Conversely, if two (sets of) variables A and B are not d-separated by C along all paths, then A and B are almost certainly statistically dependent given C, $A \not\perp\!\!\!\perp B|C$.[9]

Conditioning on a variable here refers to perfect stratification by the values of the variable. Short of perfect stratification, conditioning only on some parsimonious function of a noncollider (e.g., in a regression) may not fully block a path and could let some residual association sneak by. In practice, there are many ways by which one may condition on a variable. Most generally speaking, conditioning

[8]Occasionally, a variable may be both a collider and a common cause. In that case, conditioning on the variable may eliminate confounding bias but induce endogenous selection bias, whereas not conditioning on the variable would lead to confounding bias yet eliminate endogenous selection bias (Greenland 2003). Nevertheless, the definitions of confounding and endogenous selection remain distinct.

[9]D-Connectedness necessarily implies statistical dependence if the DAG is faithful.

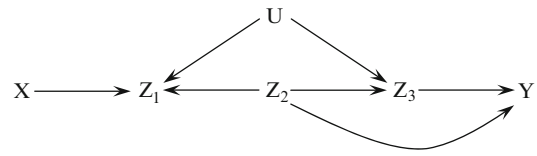

Fig. 13.6 *X* and *Y* can be d-separated and d-connected by conditioning on various sets of observed variables. *U* is unobserved

refers to incorporating some information about that variable into the analysis. This can occur in the research design stage of a study, when only individuals with certain values on variable are purposefully selected into the sample (e.g., only collecting data on employed women). Alternatively, conditioning can occur inadvertently due to attrition or nonresponse on a variable. Or it may occur because one has explicitly conditioned on a variable in the analysis stage of the study through stratification, subgroup analysis, or entering the variable as a control in a regression-type model.

Getting a firm grip on the mechanics of d-separation is essential for using DAGs. As an example, consider the DAG in Fig. 13.6, where all variables except U are observed. The variables X and Y can be d-separated, that is, rendered statistically independent, by conditioning on various sets of the observed variables. First, note that X and Y are marginally independent because all paths between X and Y contain the collider Z_1, and not conditioning on a collider blocks a path. Conditioning on Z_2 or Z_3 changes nothing, since Z_2 and Z_3 does not unblock any path between X and Y. Conditioning on Z_1 alone, however, would unblock three paths between$X$ and Y ($X \rightarrow \boxed{Z_1} \leftarrow Z_2 \rightarrow Z_3 \rightarrow Y$, $X \rightarrow \boxed{Z_1} \leftarrow U \rightarrow Z_3 \rightarrow Y$, and $X \rightarrow \boxed{Z_1} \leftarrow Z_2 \rightarrow Y$) such that X and Y would become conditionally dependent given Z_1. Conditioning on Z_1 together with Z_3 would block two of these newly opened paths, since Z_3 is a noncollider along these paths ($X \rightarrow \boxed{Z_1} \leftarrow Z_2 \rightarrow \boxed{Z_3} \rightarrow Y$ and $X \rightarrow \boxed{Z_1} \leftarrow U \rightarrow \boxed{Z_3} \rightarrow Y$), but it would not block the third open path ($X \rightarrow \boxed{Z_1} \leftarrow Z_2 \rightarrow Y$), and it would furthermore open a fourth, previously closed, path ($X \rightarrow \boxed{Z_1} \leftarrow U \rightarrow \boxed{Z_3} \leftarrow Z_2 \rightarrow Y$), since Z_3 is a collider along this path. The latter two paths could be closed again by conditioning on Z_2. In sum, X and Y are d-separated, or statistically independent, if one conditioned on either one of five sets—the empty set, Z_2, Z_3, (Z_2, Z_3), or (Z_1, Z_2, Z_3)—and X and Y would be d-connected, or statistically dependent, if one conditioned on either one of Z_1, (Z_1, Z_2), or (Z_1, Z_3). Clearly, whether two variables are dependent or independent depends not only on the structure of the data generating mechanism encoded in the DAG but also on the analyst's conditioning actions.

Although the validity of a causal assumption cannot generally be tested against the data in isolation (because that would require ruling out unmeasured confounders, which is itself a causal assumption (Robins and Wasserman 1999)), combinations of assumptions can have testable implications (see Chap. 15 by Bollen and Pearl, this volume). One of the important practical uses of d-separation is that it enumerates the *testable implications* of a causal model. Table 13.1 lists all pairwise marginal and conditional independencies implied by the DAG in Fig. 13.6.[10] The terms involving only observed variables are the empirically testable implications. The terms involving unobserved variables are not empirically testable. To the extent that the testable predictions are not substantiated in the data—subject to the usual serious caveats about type I and type II errors of significance testing—the DAG

[10]Table 13.1 assumes faithfulness. Under weak faithfulness, the DAG only implies the listed marginal and conditional independences. Many authors prefer to assume weak faithfulness. Free software for deriving testable implications is reported in Textor et al. (2011) and in Kyono (2010).

Table 13.1 All pairwise marginal and conditional independences and dependences implied by the causal DAG in Fig. 13.6

Independences		Dependences	
Marginal	Conditional	Marginal	Conditional
X and Z_2	X and Z_2 given (Z_3 or Y or U)	X and Z_1	X and Z_1 given (any other)
X and Z_3	X and Z_3 given (Z_2 or Y or U)	Z_1 and U	X and Z_2 given (Z_1 and (any other))
			X and U given (Z_1 and (any other))
X and Y	X and Z_3 given (Z_1 and Z_2 and U and (Y or ())	Z_1 and Z_2	X and Z_3 given (Z_1 and (() or (Z_2 eo U) or Y)
	X and U given (Z_2 or Z_3 or Y)		
X and U	X and Y given (U or Z_2 or Z_3)	Z_1 and Z_3	X and Y given (Z_1 and (() or (Z_2 eo (U or Z_3))))
Z_2 and U	X and Y given (Z_1 and Z_2 and (U or Z_3))	Z_1 and Y	Z_1 and U given (any others)
	Z_1 and Z_3 given (U and Z_2 and (() or X or Y))	U and Z_3	Z_1 and Z_2 given (any others)
	Z_1 and Y given (Z_2 and (U or Z_3) and (() or X))	U and Y	Z_1 and Z_3 given (X or Y or (Z_2 eo U))
	Z_2 and U given X	Z_2 and Z_3	Z_1 and U given (X or Z_3 or (U eo Z_2))
			Z_2 and U given ((Z_1 or Z_3) and (any other))
	U and Y given ((Z_2 and Z_3) and (any other))	Z_2 and Y	Z_2 and Z_3 given (any others)
			Z_2 and Y given (any others)
			U and Z_3 given (any others)
			U and Y given (X or Z_1 or (Z_2 eo Z_3))
			Z_3 and Y given (any others)

Notes: "Any other" is any combination of other variables not already named, including the empty set; "or" is the inclusive "either one or both"; "eo" is the exclusive "either one but not both"; () is the empty set

does not accurately represent the mechanism by which the data were generated. This would signal the analyst to modify her causal model. We distinguish between *weak* and *strong contradictions* between the causal model and the data. A weak contradiction occurs when a relationship that was predicted to be dependent turns out to be independent. A strong contradiction occurs when a relationship that was predicted to be independent turns out to be dependent. For example, the model would be strongly contradicted if X were not marginally independent of Z_2. Strong contradictions always imply that the DAG is incorrect. The correct DAG, however, cannot be inferred from the data alone. Among other possibilities, the correct DAG might include the arrow $X \rightarrow Z_2$, or the arrow $X \leftarrow Z_2$, or another unmeasured confounder, V, $X \leftarrow V \rightarrow Z_2$. DAGs have unambiguous implications for the independence structure of compatible probability distributions, but the independence structure of a distribution of *observed* variables is consistent with multiple DAGs. Spirtes et al. (2001) extensively discuss model testing. Pearl (2012a) discusses some problems with conventional approaches to model testing. See Robins and Wasserman (1999) and Greenland (2010) for critical perspectives. Outside of structural equation modeling, testing of model validity is extremely rare in applied sociology.

DAGs as NPSEM and Effect Heterogeneity

Causal DAGs can be read as nonparametric structural equation models (NPSEM) (Pearl 2012a), a reading that this chapter has implicitly assumed all along. Going into a little more detail helps clarify the relationship between DAGs and conventional linear path models and corrects the common misconception that DAGs cannot represent effect heterogeneity.

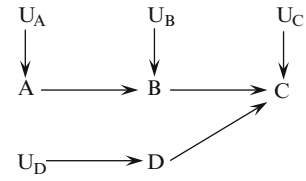

Fig. 13.7 Causal DAGs can be read as nonparametric structural equation models (*NPSEM*)

Consider the causal DAG in Fig. 13.7, which, in contrast to the usual convention, explicitly shows the independent idiosyncratic causes (error terms), U, of each variable. One can rewrite this DAG as a system of nonparametric structural equations where each endogenous variable V equals some function f_V of its parents and each U trivially equals itself:

$$A = f_A(U_A), \tag{13.1.1}$$

$$D = f_D(U_D), \tag{13.1.2}$$

$$B = f_B(A, U_B), \tag{13.1.3}$$

$$C = f_C(B, D, U_C). \tag{13.1.4}$$

These equations are nonparametric in the sense that they make no statement about distributions or functional form. The only restriction placed upon them is the structure of dependencies and independencies implied by the DAG via d-separation. (Clearly, it is easier to glean the testable implications from a DAG, for example, that D and B may be conditionally associated given C.)

Every NPSEM is consistent with a variety of parametric specifications. We say that a parametric specification is consistent with an NPSEM if it preserves the parent-child relationships and error-term independencies of the DAG. For example, the nonparametric structural equation $C = f_C(B, D, U_C)$ is consistent with the following parametric models, among others:

$$C = \alpha + \beta_B B + \beta_D D + U_C, \tag{13.2.1}$$

$$C = \beta_{B,i} B + \beta_{BD,i} \mathrm{BD} + U_C, \tag{13.2.2}$$

$$\mathrm{logit}(C) = \beta_{B,i} \sin^{-1} B + \beta_{\mathrm{BD}} \frac{\sqrt[3]{B}}{D!} + .7 * |U_C| \text{if} \quad \mathrm{DB} > 0, \quad \text{else } C = 3. \tag{13.2.3}$$

Equation 13.2.1 corresponds to a conventional linear model without effect heterogeneity that is, the parameters β_B and β_D are asserted to be the same across individuals i in the population. The others do not. This makes clear that conventional linear SEMs are highly unusual special cases of NPSEMs.

In contrast to conventional linear path models, DAGs generically presume that all causal effects vary across units unless otherwise stated. DAGs also permit effect modification and interaction effects between variables (VanderWeele and Robins 2007; VanderWeele 2009), subject only to the functional constraints embedded in the DAG. For example, the DAG in Fig. 13.7 permits that the causal effect of B on C varies with the value of D since the DAG states that C causally depends both on B and on D. This possibility is encoded in Eqs. (13.2.2) and (13.2.3) (which make different assumptions about how exactly the effect of B may vary with D). At the same time, the DAG in Fig. 13.7 rules out that the effect of D on C varies with the value of A after conditioning on B, since C does not

depend on A given B. In contrast to linear path models, where interactions are sometimes encoded with arrows pointing into arrows, effect modification and interactions in DAGs need to be read from the d-separation constraints in the model (pointing arrows into arrows would invalidate the formal syntax for working with DAGs).[11]

Estimating the statistical parameters of an NPSEM is in principle straightforward if all variables are observed and the sample is large. The chain rule of probability theory factors the joint probability distribution $P(x_1, x_2, \ldots, x_n)$ of a set of n discrete variables $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ as

$$P\left(x_1, x_2, \ldots, x_n\right) = \prod\nolimits_j P\left(x_j | x_1, \ldots, x_{j-1}\right). \tag{13.3}$$

Since every variable in a NPSEM only depends on its parents, this simplifies to

$$P\left(x_1, x_2, \ldots, x_n\right) = \prod\nolimits_j P\left(x_j | \text{pa}\left(x_j\right)\right).^{12} \tag{13.4}$$

Recovering desired causal parameters, if they are identified and if all terms in Eqs. (13.3) and (13.4) are positive, is then just a question of skillfully isolating the required parts of this distribution (as in Eq. (13.5) below). Of course, this can be difficult to impossible in practice if some variables are unobserved. The remainder of this chapter discusses various graphical identification criteria that imply valid nonparametric estimators when the causal effect is identified and the necessary variables are observed, and it discusses some challenges of inserting parametric specifications when nonparametric estimation is not feasible.

Graphical Identification Criteria

Graphical identification criteria are DAG-based rules that specify when and how identification is possible. This chapter focuses primarily on nonparametric graphical identification criteria which work regardless of how the variables are distributed and regardless of the functional form of the causal effects, that is, criteria that establish identification purely on the basis of the qualitative causal assumptions encoded in a DAG. Most graphical identification criteria discussed in the literature are sufficient (i.e., they positively determine when a causal effect can be identified), and some are necessary (i.e., they negatively determine when a causal effect cannot be identified). D-separation is the essential tool underlying all graphical identification criteria.

Identification by Covariate Adjustment

The Adjustment Criterion

Most empirical approaches to causal inference rely on adjusting (conditioning, controlling, stratifying) for numerous covariates (the so-called adjustment set) in some kind of regression model in order to strip an observed association of all spurious components. The danger of this strategy is that

[11] Morgan and Winship (2012) use DAGs to discuss effect heterogeneity in sociological applications. Elwert and Winship (2010) use DAGs to show that unobserved effect heterogeneity can bias linear regression models.

[12] This equation is the causal Markov condition, mentioned earlier.

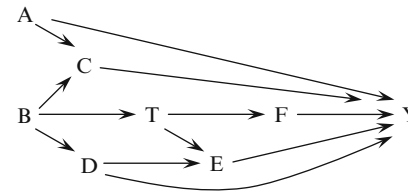

Fig. 13.8 Multiple adjustment sets satisfy the adjustment criterion relative to the total causal effect of T on Y

unprincipled covariate adjustment may fail to remove all confounding bias or even introduce new biases through overcontrol or endogenous selection. Various graphical identification criteria exist to guide the proper choice of covariates if a sufficient set of covariates is in fact observed.

The main insight of the graphical approach to covariate adjustment is that the adjustment set must block all noncausal paths without blocking any causal paths between treatment and outcome. This is accomplished if the adjustment set meets the adjustment criterion.

Adjustment criterion (Shpitser et al. 2010): A set of observed variables Z (which may be empty) satisfies the adjustment criterion relative to the total causal effect of a treatment T on an outcome Y if:

1. Z blocks all noncausal paths from T to Y; and
2. No variable in Z lies on or descends from a causal path from T to Y.[13]

If Z satisfies the adjustment criterion, then the total causal effect of T on Y is nonparametrically identified by adjusting for Z (Shpitser et al. 2010).

Figure 13.8 illustrates the adjustment criterion relative to the total causal effect of T on Y. T and Y are connected by nine distinct paths. Of these, two are causal paths, $T \rightarrow F \rightarrow Y$ and $T \rightarrow E \rightarrow Y$, which together comprise the total causal effect of interest. The other seven paths are noncausal paths that need to be blocked. Four of the noncausal paths are already (unconditionally) blocked because they contain colliders: $T \leftarrow B \rightarrow C \leftarrow A \rightarrow Y$, where C is the blocking collider; $T \rightarrow E \leftarrow D \rightarrow Y$, where E is the blocking collider; and $T \rightarrow E \leftarrow D \leftarrow B \rightarrow C \rightarrow Y$ and $T \rightarrow E \leftarrow D \leftarrow B \rightarrow C \leftarrow A \rightarrow Y$, where E and C are blocking colliders. The remaining three noncausal paths are open, but they can be blocked by conditioning on B: $T \leftarrow \boxed{B} \rightarrow C \rightarrow Y$, $T \leftarrow \boxed{B} \rightarrow D \rightarrow E \rightarrow Y$, and $T \leftarrow \boxed{B} \rightarrow D \rightarrow Y$. Since B does not lie on or descend from a causal path, the adjustment criterion for the total causal effect of T on Y is met by conditioning on the adjustment set $Z = B$.

More than one adjustment set may meet the adjustment criterion. Instead of conditioning on B, we could block two of the three open noncausal paths by conditioning on D: $T \leftarrow B \rightarrow \boxed{D} \rightarrow E \rightarrow Y$ and $T \leftarrow B \rightarrow \boxed{D} \rightarrow Y$. And we could block the third open noncausal path, $T \leftarrow B \rightarrow \boxed{C} \rightarrow Y$, by conditioning on C. The last move, however, is problematic since conditioning on C opens up the previously closed noncausal path $T \leftarrow B \rightarrow \boxed{C} \leftarrow A \rightarrow Y$ on which C is a collider. This newly opened noncausal path can be blocked again by conditioning on A. Thus, conditioning on $Z = (A, C, D)$ would satisfy the adjustment criterion.

This example demonstrates several interesting facts about identification by adjustment. First, a given causal effect can often be identified by multiple possible adjustment sets. In Fig. 13.8, there are nine possible adjustment sets for the total causal effect of T on Y: (B), (B, A), (B, C), (B, D), (B, A, C), (B, C, D), (B, A, D), (A, C, D), and (A, B, C, D). Either one of these adjustment sets identifies the causal effect well as the others. Second, it may not be necessary to condition on any direct causes of treatment (the so-called assignment mechanism)—conditioning on (A, C, D) will do. Third, it may not

[13] The requirement not to condition on a descendant of a variable on a causal path is explained in the discussion of Fig. 13.9 below.

be necessary to condition on any direct causes of the outcome—conditioning on B will do. Fourth, it may not be necessary to condition on any joint ancestors of treatment and outcome—conditioning on (A, C, D) will do. These facts grant the analyst considerable flexibility in the identification of a causal effect, which is useful if one or more of the covariates are unobserved, such that one or more of the possible adjustment strategies are rendered infeasible.

Adjustment and Ignorability

The adjustment criterion is theoretically important because it provides a direct link to the potential outcomes framework of causal inference. Conditional ignorability of treatment assignment with respect to the potential outcomes, $Y^T \perp\!\!\!\perp T|Z$, (Rosenbaum and Rubin 1983) implies that the adjustment criterion is met by adjusting for Z; and the fact that the adjustment criterion is met implies conditional ignorability (Shpitser et al. 2010). Knowledge of the DAG therefore helps analysts understand when conditional ignorability is met and when it is not met. Best of all, it does so by making reference only to qualitative causal statements about in-principle observable variables encoded in the DAG, not always-unobservable counterfactuals.

Estimation When the Adjustment Criterion Is Met

The correspondence between the adjustment criterion and conditional ignorability, together with elementary rules of probability theory, gives rise to a straightforward nonparametric estimator. Write $P(v)$ for the probability distribution of V, $V = v$. If Z meets the adjustment criterion for the total causal effect of T on Y, then the distribution of the potential outcomes can be estimated nonparametrically from observed data (Z, T, Y) by

$$P\left(y^T\right) = \sum\nolimits_z P\left(y|t,z\right) P(z) \tag{13.5}$$

or its continuous analogue (Robins 1986; Pearl 1995; Shpitser et al. 2010).[14] Equation (13.5) is known by various names, including stratification estimator and, lately, adjustment formula. For a binary treatment and categorical Z, the total effect of T on Y evaluates to

$$\begin{aligned} &E\left[Y^{T=1}\right] - E\left[Y^{T=0}\right] \\ &\quad = \sum\nolimits_z E\left[Y|T=1, Z=z\right] \Pr(Z=z) - \sum\nolimits_z E\left[Y|T=0, Z=z\right] \Pr(Z=z). \end{aligned} \tag{13.6}$$

Nonparametric estimators for arbitrarily distributed variables follow analogously.

Unfortunately, the nonparametric estimator in Eq. (13.5) is rarely feasible in practice. If Z is high dimensional or the sample is small, the analyst may need to insert parametric functions into the terms of Eq. (13.5). It is important to realize that things can go badly wrong at this step. Just because a causal effect is proved nonparametrically identifiable by the adjustment criterion does not imply that it can be estimated with just any parametric estimator (e.g., by throwing the variables in the adjustment set, $Z = Z_1, \ldots, Z_n$, as main effects into a linear regression $Y = a + b_T T + \sum_k b_k Z_k + e$). The

[14] Pearl (1995, 2009) and others use so-called *do-operator notation* to write $P\left(y^T\right)$ as $P\left(Y = y \,|\, \mathrm{do}(T = t)\right)$. The do-operator $do(T = t)$ emphasizes that T is set to t by intervention ("doing"). $P(Y = y \,|\, \mathrm{do}(T = t))$ gives the post-intervention distribution of Y if one intervened on T to set it to some specific value t, that is, the counterfactual distribution of T.

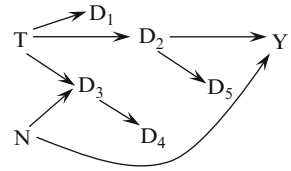

Fig. 13.9 A DAG for a randomized experiment that illustrates that conditioning on descendants of the treatment is either irrelevant or harmful for identification by adjustment

appropriateness of a specific parametric estimator depends on the appropriateness of its parametric assumptions for the specific substantive situation at hand, about which DAGs typically have little to say. Breen and Karlson (Chap. 10, this volume) discuss an example where the causal effect of a treatment on a binary outcome is nonparametrically identified by the adjustment criterion (and thus could be estimated nonparametrically using Eq. (13.5)), and yet the parametric assumptions embedded in off-the-shelf logit and probit models induce tricky biases. The problem in such cases lies not with identification, but with parametric estimators that have introduced faulty parametric assumptions.

Sufficient Conditions for the Adjustment Criterion

The adjustment criterion implies several narrower identification criteria that offer useful guidance for quickly finding sufficient adjustment sets in practice.

First, note that it is never necessary, and often harmful, to condition on a descendant of the treatment. Consider the total causal effect of T on Y in Fig. 13.9, which is identified unconditionally since T is randomized. Conditioning on a descendant of T that lies on or descends from the causal path is always harmful because it either controls away the effect of interest (D_2) or induces endogenous selection bias (D_5).[15] Conditioning on a descendant of T that lies on (D_3), or descends from (D_4), a noncausal path is never necessary because such paths are by definition blocked by a collider, and it may be harmful as it may unblock the path. Conditioning on a descendant of T that neither lies nor descends from a causal or noncausal pathway (D_1) is neither necessary nor harmful to identification (but it will reduce efficiency). Analysts interested in identifying the total causal effect of a treatment are thus well advised not to adjust for a descendant of treatment.

The analyst therefore only needs to worry about noncausal paths that start with an arrow into treatment, $\rightarrow T$. Such noncausal paths are called *backdoor paths*.

Backdoor criterion (Pearl 1993, 2009): A set of observed variables Z (which may be empty) satisfies the backdoor criterion relative to the total causal effect of a treatment T on an outcome Y if:

1. No element of Z is a descendant of T; and
2. Z blocks all backdoor paths from T to Y.

If Z satisfies the backdoor criterion, then the total causal effect of T on Y is nonparametrically identified (Pearl 1995). In Fig. 13.10, the backdoor criterion is met by seven distinct adjustment sets: (A, C), (B, C), (C, D), (A, B, C), (A, C, D), (B, C, D), and (A, B, C, D). All effects identified by the adjustment criterion are also identified by the backdoor criterion (Shpitser et al. 2010).

[15] D_5 is a descendant of the collider $T \rightarrow D_2 \leftarrow e_2$ (recall the implied existence of idiosyncratic error terms), which opens the noncausal path T- - - $e_2 \rightarrow D_2 \rightarrow Y$.

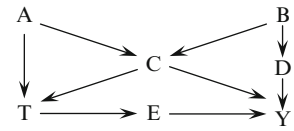

Fig. 13.10 Illustrating various narrower identification criteria implied by the adjustment criterion

Sometimes, identification can be detected at a glance. Since all backdoor paths by definition start with an arrow into treatment, all backdoor paths can be blocked by conditioning on the parents of treatment, if they are observed.

Parents of treatment criterion (Pearl 1995): Conditioning on all parents of treatment, T, identifies the total causal effect of T on any outcome.

In Fig.13.10, the parents of treatment criterion is met by conditioning on (A, C). By contrast, conditioning on all parents of the outcome might induce overcontrol bias since the parents of the outcome usually include variables on a causal pathway between T and Y. However, under certain circumstances, restricting adjustment to parents that do not lie on a causal pathway is sufficient.

Parents of the outcome criterion: If no backdoor path shares a node with any causal path (other than T and Y), then conditioning on all parents of the outcome Y that do not lie on a causal path from T to Y identifies the total causal effect of T on Y.

In Fig. 13.10, the parents of the outcome criterion is met by conditioning on (C, D), but not by conditioning on (C, D, E). Next, note that all unconditionally open backdoor paths must contain a variable that is a joint direct or indirect cause of treatment and outcome, $T \leftarrow \cdots \leftarrow J \rightarrow \cdots \rightarrow Y$. The following criterion can then be shown.

Joint ancestor criterion: Conditioning exclusively on all joint ancestors of T and Y identifies the total causal effect of T on Y. (Conditioning on additional variables may ruin identification.)

In Fig. 13.10, the joint ancestor criterion is met by conditioning on (A, B, C). Of course, all of these graphical identification criteria will only work if at least one possible adjustment set is actually observed. Conditioning only on the *observed* parents of treatment, or the *observed* joint ancestors of treatment and outcome, may fail to remove all bias or even create new biases. If only T, Y, and C were observed in Fig. 13.10, then conditioning on C as the sole observed parent of T, parent of Y, or joint ancestor of T and Y will not identify the causal effect (since conditioning on C opens a noncausal path $T \leftarrow A \rightarrow \boxed{C} \leftarrow B \rightarrow D \rightarrow Y$). In practice, it is advisable to use the backdoor criterion, since the backdoor criterion may detect a possible adjustment set even if none of the narrower parent or ancestor criteria are met.

Finally, the backdoor criterion implies a very helpful identification criterion that works (i.e., avoids inadvertent bias) even if the structure of the DAG is not fully known.

Confounder selection criterion (VanderWeele and Shpitser 2011): If there is a set of observed covariates that meets the backdoor criterion (i.e., if the analyst is willing to assume ignorability), then it is sufficient to condition on all observed pretreatment covariates that either cause treatment, outcome, or both.

If the total causal effect of T on Y is identified by any of the above criteria, then it can be estimated nonparametrically by Eq. (13.5).

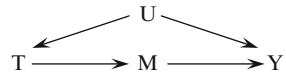

Fig. 13.11 The total effect of T on Y is identifiable via the frontdoor criterion but not via the adjustment criterion. U is unobserved

Identification Beyond Adjustment: Frontdoor Identification, the do-Calculus, and Instrumental Variables

Covariate adjustment is not the only road to nonparametric identification. Numerous additional strategies exist that may work even if treatment is not ignorable and no set of covariates satisfies the adjustment criterion. All of these strategies can be represented graphically with DAGs, and all ultimately rely on d-separation.

One of the alternatives to the adjustment criterion is Pearl's (1995) *frontdoor identification criterion*. Frontdoor identification relies on piecing together a total causal effect from its constituent parts through repeated application of the backdoor criterion. For example, the total causal effect of T on Y in Fig. 13.11, where U is unobserved, is not identifiable via the adjustment criterion because the backdoor path $T \leftarrow U \rightarrow Y$ cannot be blocked. All segments of the causal path, $T \rightarrow M \rightarrow Y$, however, can be identified separately. $T \rightarrow M$ is identified by the marginal association between T and M (because no open backdoor path connects them); and $M \rightarrow Y$ is identified by the conditional association between M and Y given T (since T lies on the only open backdoor path). Piecing the parts together is a matter of straightforward algebra. Pearl (2009: 83) gives a nonparametric estimator for frontdoor-identified causal effects. See Knight and Winship (Chap. 14, this volume) for a detailed discussion of frontdoor identification.

The most general nonparametric graphical identification criterion is Pearl's *calculus of intervention*, or *do-calculus* (1995, 2009: Section 3.4). Shpitser and Pearl (2006) prove that the do-calculus is complete in that it detects the identifiability of all nonparametrically identifiable causal effects of interventions. All other graphical identification criteria, including the adjustment (ignorability) and frontdoor criteria, are special cases of the do-calculus.

Few of the nonparametric identification strategies beyond covariate adjustment are known outside of specialist circles. Many of these strategies make exacting demands on data and theory by requiring clever exclusion restrictions. Careful development of sociological theory will no doubt reveal fresh opportunities to apply these more advanced graphical identification criteria for empirical gain.

If a causal effect is not identifiable by any of the above nonparametric criteria (e.g., if treatment and a child of treatment on a causal pathway are confounded by an unobserved variable), then the analyst may be still be able to achieve identification if he or she can defend additional parametric assumptions. Pearl (2009: chapter 5) discusses graphical criteria for identification in linear models; Brito and Pearl (2002) give a graphical criterion for detecting instrumental variables and instrumental sets in DAGs; and Chan and Kuroki (2010) give graphical criteria for instrument-like auxiliary variable identification strategies.

The Importance of Having a Causal Model: Confounding as a Causal Rather Than an Associational Concept

The graphical identification criteria reviewed above all presume the validity of the causal model encoded in the DAG. This comports with the claim that identification analysis requires an explicit

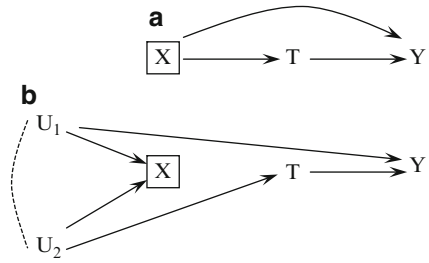

Fig. 13.12 *X* fits the associational definition of confounding in both DAGs. But conditioning on *X* in (**a**) removes bias, whereas in (**b**) it creates bias. U_1 and U_2 are unobserved

causal model. I will now draw on recent work on the nature of confounding to prove that having a causal model is in fact necessary and that purely observational, atheoretical approaches can lead the analyst astray (Greenland and Robins 1986; Greenland et al. 1999b; Cole and Hernán 2002).

Consider a conventional approach to confounder selection. Textbooks often define a confounder as a variable that (1) temporally precedes treatment and is both (2) *associated* with the treatment and (3) *associated* with the outcome. Standard recommendations dictate that such variables must be controlled to avoid bias.

The two DAGs in Fig. 13.12 show why this associational rule can lead to the introduction rather than removal of bias. In both DAGs, *X* fulfills the three conventional criteria of a confounder for the causal effect of *T* on *Y*: *X* precedes *T*, *X* is associated with *T*, and *X* is associated with *Y* (by direct causation in Fig. 13.12a and via unobserved variables *U* in Fig. 13.12b). The conventional prescription works in Fig. 13.12a: *X* sits on the only open noncausal path from *T* to *Y* and must be conditioned on to eliminate bias. By contrast, the conventional prescription fails in Fig. 13.12b: There is no open noncausal path between *T* and *Y* because *X* is a collider. Controlling for *X* would open this noncausal path and induce endogenous selection bias. The problem is compounded because there is no way to distinguish empirically between Fig. 13.12a, b, as both DAGs have the same set of testable implications (all observed variables are conditionally and unconditionally associated with each other). The lessons of this example are that it is not possible to decide on the proper set of control variables without an understanding of the underlying causal structure and that observable, associational, criteria alone cannot justify the identification of causal effects.

Examples of Endogenous Selection Bias

One of the practical uses of DAGs lies in elucidating the nature of selection bias as conditioning on a collider (Hernán et al. 2004; Elwert and Winship forthcoming). Selection bias can occur at any time in a causal process. It can occur due to conditioning on an outcome or a descendant of an outcome; it can occur due to conditioning on an intermediate variable affected by the treatment; and it can even occur due to conditioning on a pretreatment variable (Greenland 2003).

Heckman Selection and the Motherhood Wage Penalty: Conditioning on a Descendant of the Outcome Creates Bias

So-called Heckman sample selection bias can be explicated as bias arising from conditioning on a post-outcome collider (Elwert and Winship forthcoming). Suppose that we are interested in estimating the total causal effect of having a child, *M*, on the wages offered to women by potential employers,

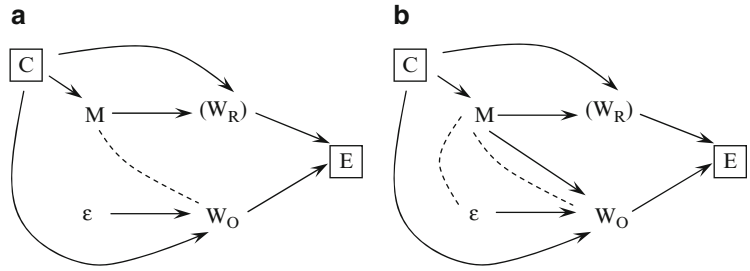

Fig. 13.13 Conditioning on a post-outcome collider can induce multiple endogenous selection biases. *M*, motherhood; W_R, reservation wage (unmeasured); W_O, offer wage; *E*, employment status; ε, "error" term on offer wages; *C*, common causes of motherhood and wages. (**a**) No effect of motherhood on offer wages. (**b**) With effect of motherhood on offer wages

W_O, (Fig. 13.13a). We assume that motherhood will affect a woman's reservation wage, W_R (i.e., the wage that would be necessary to draw her into the workforce; see Gronau 1974; Heckman 1974). W_R is not measured. The decision to accept employment, *E*, is affected both by the offer wage, W_O, and the reservation wage, W_R, because a woman will only accept the job if the offer wage meets or exceeds her reservation wage. Employment thus is a collider on a noncausal path between motherhood and offer wages, $M \rightarrow W_R \rightarrow E \leftarrow W_O$. A vector of common causes, *C*, confounds these relationships, but, for simplicity, we assume that the analyst has appropriately conditioned on them to block all backdoor paths between *M* and W_O.

The central problem is that most datasets include information on offer wages only for those women who are actually employed. Many analyses have therefore restricted the sample to employed women. But this sample restriction implies conditioning on the collider *E*, which unblocks the noncausal path from motherhood to offer wages, $M \rightarrow W_R \rightarrow \boxed{E} \leftarrow W_O$, and induces endogenous selection bias: The analysis would detect an association between motherhood and wages even if the causal effect of motherhood on wages were in fact zero (Fig. 13.13a).

If motherhood has no effect on offer wages, as assumed in Fig. 13.13a, the endogenous selection problem is bad enough. A second source of endogenous selection bias is introduced if motherhood really has an effect on offer wages (e.g., because of mothers' differential productivity or employer discrimination). This is shown in Fig. 13.13b by adding the arrow $M \rightarrow W_O$. Since all outcomes are implicit colliders between the treatment and the error term (if there is an effect of treatment), conditioning on *E* now also amounts to conditioning on the descendant of the collider W_O along the path $M \rightarrow W_O \leftarrow \varepsilon$, which induces a new noncausal association between motherhood and the error term, and from there with offer wages, $M \dashrightarrow \varepsilon \rightarrow W_O$. Note that this second endogenous selection problem (but not the first) would exist even if we could measure and condition on W_R.[16]

With a few additional assumptions, one can predict the direction of the bias (VanderWeele and Robins 2009). If motherhood decreases the chances of employment and if higher offer wages increase the chances of accepting employment for all women, then mothers who are employed must on average have received higher offer wages than childless women. Consequently, an analysis that is restricted to working women would underestimate the motherhood wage penalty.

[16]The difference between Fig. 13.13a, b illustrates why identifying the magnitude of a causal effect is more difficult than testing the null of no effect. If one could condition on W_R in Fig. 13.13a, then the absence of an association between *M* and W_0 conditional on *E* and W_R would imply the absence of a causal effect $M \rightarrow W_0$—the null can be tested. But if there is an effect $M \rightarrow W_0$, as in Fig. 13.13b, then the observed association between *M* and W_0 given *E* and W_R is biased for the causal effect $M \rightarrow W_0$—the magnitude of the effect cannot be measured.

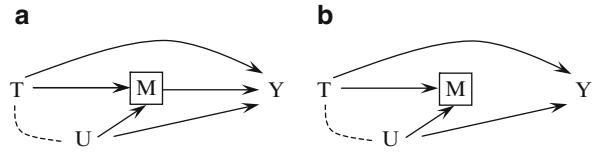

Fig. 13.14 *T*, randomized treatment; *M*, nonrandomized variable affected by *T*; *Y*, outcome; *U*, unobserved common cause. (**a**) Conditioning on the mediator *M* induces endogenous selection that biases the regression estimate for the direct effect of *T* on *Y*. (**b**) The marginal association between *T* and *Y* does not equal the conditional association between *T* and *Y* given *M*, falsely suggesting the presence of an "indirect" effect via *M* where no such indirect effect exists

Direct Effects, Indirect Effects, and Mediation Analysis: Conditioning on an Intermediate Variable Can Create Bias

Outside of certain unlikely scenarios, conventional mediation analysis (e.g., following Baron and Kenny 1986), and with it the estimation of causal mechanisms, direct and indirect effects, is virtually guaranteed to suffer endogenous selection bias. The problem is well known in methodological circles (e.g., Rosenbaum 1984; Holland 1988; Robins 1989; Smith 1990; Wooldridge 2005, 2006; Sobel 2008), but it stubbornly persists in empirical social science. DAGs readily communicate the essence of the problem as endogenous selection bias (Pearl 1998; Robins 2001; Cole and Hernán 2002).

To fix ideas, consider whether class size in first grade, *T*, has a direct effect on high school graduation, *Y*, $T \rightarrow Y$, via some mechanism other than boosting student achievement in third grade, *M* (Finn et al. 2005). Assume that class size is randomized, as in the well-known Project STAR experiment (Finn and Achilles 1990). Figure 13.14a gives the basic corresponding DAG. As in any well-executed randomized experiment, the total effect of treatment, *T*, on the outcome, *Y*, is identified by the marginal association between *T* and *Y* because *T* and *Y* share no common cause. As in all observational studies, however, the posttreatment mediator, *M*, is not randomized and may therefore share an unmeasured cause, *U*, with the outcome, *Y*. Candidates for *U* in this application might include parental education, underlying ability, student motivation, and any other confounders of *M* and *Y* not explicitly controlled in the study. The existence of the confounder *U* would make *M* a collider variable. Conditioning on *M* in order to estimate the direct causal effect $T \rightarrow Y$ would unblock the noncausal path $T \rightarrow \boxed{M} \leftarrow U \rightarrow Y$ and induce endogenous selection bias. Therefore, the direct effect $T \rightarrow Y$ in this DAG is not identified by conditioning on *M*.

The conventional strategy for detecting *indirect* effects is similarly susceptible to endogenous selection bias. It is a common practice to infer the existence of an indirect effect of *T* on *Y* by comparing an estimate for the total effect of *T* on *Y* with an estimate for the direct effect of *T* on *Y*. If the total effect estimate (e.g., from a regression of *Y* on *T*) differs from the naïve direct effect estimate (e.g., from a regression of *Y* on *T* and *M*), analysts commonly conclude that there must be an indirect effect and that *M* is a "mediator." Figure 13.14b shows why this strategy may lead to the wrong conclusion. In this DAG, the total effect of *T* on *Y* is identical with the direct effect because no indirect effect of *T* on *Y* via *M* exists. The total effect is identified by the marginal association between *T* and *Y*. The conditional association between *T* and *Y* given *M*, however, will differ from the marginal association between *T* and *Y* because *M* is a collider, and conditioning on the collider induces endogenous selection bias along the noncausal path $T \rightarrow \boxed{M} \leftarrow U \rightarrow Y$. Thus, the correct estimate for the total causal effect and the biased estimate for the direct effect of *T* on *Y* would differ, and the analyst would falsely conclude that an indirect effect exists even though it does not.

Generally, neither direct nor indirect effects are nonparametrically identifiable by simple conditioning on the mediator if there exist unmeasured common causes of the mediator and the outcome (but see the section on time-varying treatments below).

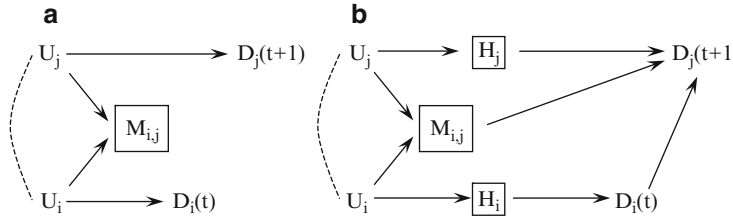

Fig. 13.15 Homophily in social network analysis is endogenous selection bias. M_{ij}, marital status of woman i and man j; D, vital status; U, characteristics influencing marital choice and vital status; H, health in old age. (**a**) Computing the association between D_i and D_j implies conditioning on M_{ij}, which induces an association between D_i and D_j even if D_i exerts no causal effect on D_j. (**b**) If D_i affects D_j only if i and j are married (effect modification), then the existence of the effect implies *two* arrows in the DAG, $D_i \rightarrow D_j$ and $M_{ij} \rightarrow D_j$. Conditioning on either one of H_i or H_j would block the noncausal path opened by conditioning on the social tie M_{ij} and allow for the identification of the causal effect of D_i on D_j

Homophily in Social Network Analysis: Conditioning on Pretreatment Variables Can Create Bias

One of the more surprising results of recent research on causal inference is that controlling for certain pretreatment variables—colliders—can increase rather than decrease bias (Pearl 1995, 2009; Greenland et al. 1999a, b; Hernán et al. 2002, 2004; Greenland 2003).[17] Even more surprising is that a central problem of modern social network analysis—latent homophily bias—is exactly of this type (Shalizi and Thomas 2011).

Consider one of the simplest social networks—the marital dyad—and ask whether the death of a wife i at time t, D_i, exerts a causal effect on the subsequent death of her husband j (Fig. 13.15). It has long been known that spouses tend to die in short succession. It has also long been suspected that this association may be owed not to a causal effect of one death on the other, but to spousal similarity, as like marries like (homophily) (Farr 1858). With the appropriate DAG, we can see that homophily bias is best understood as endogenous selection rather than confounding (Elwert and Winship forthcoming). Consider Fig. 13.15a, which encodes the null hypothesis that wife's death does not affect husband's mortality. It further assumes—for expositional clarity—that husband's and wife's vital statuses are not confounded by any common cause. However, there are unobserved factors, U_j, such as husband's education, that affect both his decision of whom to marry (and to stay married to), M_{ij}, and his vital status, D_j. Similarly, wife's unobserved education, U_i, affects both her decision to marry this specific husband and her vital status. If this DAG is true, then there is no open noncausal path between D_i and D_j, and husband's and wife's vital status should be marginally independent. The trouble is that one cannot observe this marginal association among married couples, because the simple act of searching for an association between the vital statuses of husbands and wives means that the analyst is conditioning on marital status, $M_{ij} = 1$. Conditional on marital status, U_i and U_j become associated: Knowing that a man and a woman are married to each other permits us to infer something about her education from his education. (If he has high education, then likely so does she.) Ultimately, conditioning on M_{ij} induces an association between D_i and D_j along the path $D_i \leftarrow U_i \rightarrow \boxed{M_{ij}} \leftarrow U_j \rightarrow D_j$. Therefore, wives' deaths will be observationally associated with husbands' deaths even if one does not cause the other.

[17] Propensity score analysis is not immune to this problem (Shrier 2009; Sjölander 2009).

This homophily problem is pervasive in network analysis, since social ties between spouses, friends, and any other kind of network alters are almost never formed at random. If tie formation or tie dissolution are affected by unobserved variables that are, respectively, associated with the treatment variable in one individual and the outcome variable in the other individual, then searching for interpersonal effects will induce a spurious association between individuals in the network.

Nevertheless, like in any other setting, causal inference in social networks is possible if the biasing paths can be blocked (Shalizi and Thomas 2011). For the present example, Elwert and Christakis (2006) argue in essence that the effect of husbands' and wives' U (e.g., unobserved education) on vital status should be substantially mediated by their health (and other observables) in old age, H. Conditioning on good measures of health would thus permit the identification of interspousal health effects (Fig. 13.15b).[18]

Greenland (2003) provides rules of thumb for the size of pretreatment endogenous selection bias. VanderWeele (2011) develops a formal sensitivity analysis. O'Malley et al. (2012) use DAGs to justify instrument-variable solutions for homophily bias. Fowler and Christakis (2010) avoid homophily bias altogether by experimentally randomizing network structure to eliminate the influence of all possible observed and unobserved variables on tie formation.

Drawing DAGs for Social Networks

The previous example shows that DAGs can inform causal inference in social networks.[19] The mechanics of graphical analysis are the same in social networks as elsewhere, but the structure of DAGs for social networks merits some comments. First, DAGs for social networks should include the social ties between individuals (or groups of individuals) as variables in their own right, for example, M_{ij}. Second, the DAG should contain separate variables for the attributes and actions of each individual (or groups of individuals) in the network, for example, D_i and D_j.[20] Third, the DAG should explicitly include the mechanism of tie formation, noting that tie formation is usually influenced by the attributes and actions of all individuals linked by the tie, for example, $U_i \rightarrow M_{ij} \leftarrow U_j$. Fourth, if one individual causally influences another individual, then this implies not only a causal path from the treatment to the outcome but also a direct arrow from the social tie into the outcome (Shalizi and Thomas 2011). In our example, suppose that Ingrid's death, D_i, increases the risk of Jack's death, D_j, only if Ingrid and Jack are married, $M_{ij} = 1$, but not if they are strangers, $M_{ij} = 0$. This effect modification of the effect of D_i on D_j by the value of M_{ij} suggests that the D_j causally depends on D_i and M_{ij}, such that the DAG should contain *two* arrows, $D_i \rightarrow D_j$ and $M_{ij} \rightarrow D_j$, to represent the causal effect of D_i on D_j (Fig. 13.15b). Fifth, investigating the spread of attitudes, states, or behaviors along a social tie *necessarily* implies that the analyst is conditioning on the tie, which is a collider, and hence risks inducing endogenous selection bias.

[18] Elwert and Christakis (2008) use additional knowledge of the network topology to gage and remove the bias from residual confounding (i.e., if conditioning on H does not solve the problem).

[19] By definition, interpersonal causal effects in social networks violate the no-interference decree of Rubin's (1980) stable unit treatment value assumption (SUTVA). See VanderWeele and An (Chap. 17, this volume) for a detailed discussion of causal inference with interference.

[20] DAGs for triadic networks would usually include separate variables for the characteristics of all three members of a generic triad. Obviously, the complexity of a DAG increases with the complexity of social structure. This is one reason why causal inference in social networks is a difficult problem.

The Sequential Backdoor Criterion for Time-Varying Treatments

The identification and estimation of causal effects from multiple, time-varying interventions has been one of the most exciting areas of causal inference over the past 15 years. In contrast to work on single, fixed-time interventions, however, the fruits of this literature are only slowly making inroads in applied social science research. In this section, I review the sequential backdoor criterion for the joint causal effect of time-varying treatments (Pearl and Robins 1995).

To fix ideas, suppose that we are interested in the joint causal effect of taking specific sequences of courses during the fall and spring terms of junior year of high school, A_0 and A_1, on a student's SAT score during senior year, Y. Suppose that students must choose between two subjects in each term t, $A_t = 0$ (math) or $A_t = 1$ (English). Students' time-varying GPA, L_t, is affected by past course choice, and L_t in turn affects future course choice as well as SAT scores. Students' unobserved test-taking ability, U, affects their GPA and their SAT score, but not their course choice. Figure 13.16a shows the DAG for this causal model.

Joint causal effects are causal effects of multiple and possibly time-varying interventions (so-called unit treatments); they are defined as the change in the outcome that would occur if one intervened to change all unit treatments from one level to another. For example, we might be interested in the joint causal effect on SAT scores of taking the sequence (math, math) rather than the sequence (English, English).[21] In a DAG, the joint causal effect of multiple interventions is represented by all those causal paths emanating from the unit treatments to the outcome that are not mediated by later unit treatments (so-called proper causal paths (Shpitser et al. 2010)). In Fig. 13.16a, the joint causal effect of changing A_0 and A_1 is captured by the three proper causal paths $A_0 \to Y$, $A_0 \to L_1 \to Y$, and $A_1 \to Y$. Note that the path $A_0 \to L_1 \to A_1 \to Y$ is not part of the desired joint causal effect because it is not a proper causal path: Intervening on A_0 and A_1 prevents A_0 from affecting A_1, thus rendering the path inactive.[22]

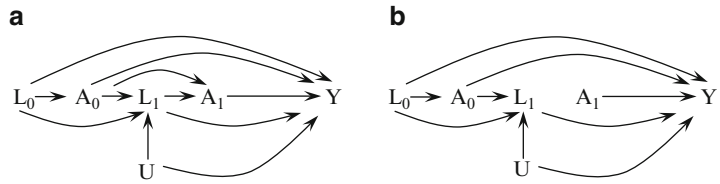

Fig. 13.16 A_t, course choice in term t; L_t, GPA going into t; U, test-taking ability; Y, SAT score. (**a**) DAG describing a causal model for the relationship between course choice and SAT score. (**b**) Modified DAG drawn from the perspective of A_0, omitting all arrows into the later unit treatment A_1. All causal paths from A_0 to Y in this redrawn DAG are proper causal paths

[21] Here, we focus on causal effects of time-varying treatments that contrast predetermined treatment sequences. For two binary unit treatments, we can define six causal effects corresponding to the six pairwise contrasts between the four possible predetermined treatment sequences, here, (math, math), (math, English), (English, math), and (English, English). Note that some of these causal effects, such as (math, English) vs. (English, English), equal so-called *controlled direct effects* (Pearl 2001; Robins and Greenland 1992). The identification criteria discussed in this section apply to all causal effects of predetermined treatment sequences and hence to all controlled direct causal effects. See Bollen and Pearl (Chap. 15, this volume) and Wang and Sobel (Chap. 12, this volume) for mediation formulae and the identification of other types of ("natural" or "pure") direct and indirect effects. See Robins and Richardson (2011) and Pearl (2012b) for graphical identification conditions of path-specific effects. See Robins and Hernán (2009) for yet other types of time-varying treatments, especially the distinction between static and dynamic time-varying treatment effects.

[22] Note that the joint causal effect of A_0 and A_1 is not the same as the total causal effect of A_0 plus the total causal effect of A_1, as is sometimes incorrectly thought.

Realizing that the joint causal effect only comprises proper causal paths simplifies identification analysis tremendously because it compartmentalizes the task: Each proper causal path belongs unambiguously to a specific unit treatment ($A_0 \rightarrow Y$ and $A_0 \rightarrow L_1 \rightarrow Y$ belong to A_0; and $A_1 \rightarrow Y$ belongs to A_1). We can thus ascertain separately for each unit treatment whether its proper causal paths can be identified. If all proper causal paths of every unit treatment can be identified, then the joint causal effects of the unit treatment can be identified.

Graphically, this is assessed by first redrawing the DAG multiple times, once each from the perspective of each specific unit treatment, A_t, such that the redrawn DAG contains only the proper causal paths belonging to A_t (in addition to all noncausal paths). This boils down to deleting all arrows into future unit treatments, $A_{t+1}, A_{t+2} \ldots, A_n$, downstream from A_t. The redrawn DAG for the proper causal effects of A_0 in our example is given in Fig. 13.16b, which omits the arrows $A_0 \rightarrow A_1$ and $L_1 \rightarrow A_1$. The redrawn DAG for A_1 equals the original DAG in Fig. 13.16a, because A_1 is the last unit treatment and hence there are no arrows into later unit treatments to be deleted.

Next, we apply the backdoor criterion to each redrawn DAG, starting with the first unit treatment and sequentially progressing to later unit treatments, to check if the total causal effect of each unit treatment on the outcome can be identified. With a few more technical details, this procedure gives the following sufficient graphical identification criterion:

Sequential backdoor criterion for the causal effect of intervening on all A_t on Y (Pearl and Robins 1995):

1. Begin with the first unit treatment A_t, $t = 0$.
2. Redraw the DAG for A_t by deleting from the original DAG all arrows into future unit treatments $A_{t+1}, A_{t+2}, \ldots, A_n$. Check if the total causal effect of A_t on Y in the redrawn DAG can be identified by the backdoor criterion. If so, select a minimally sufficient set of covariates that meet the backdoor criterion and call it Z_t[23]; then repeat step 2 for the next unit treatment, $t = t + 1$. If not, the joint causal effect is not identified by this criterion.
3. If step 2 succeeds for all unit treatments, then the joint causal effect of intervening on all A_t on Y is identified.

This criterion is less complicated than it sounds. In our example, one would begin by investigating the effect of the first unit treatment, A_0 on Y in the redrawn DAG that omits all arrows into the sole future unit treatment, A_1 (Fig. 13.16b). Applying the backdoor criterion to this redrawn DAG, we see that the total causal effect of A_0 on Y is identified by conditioning on $Z_0 = L_0$, since L_0 blocks all backdoor paths. Then, move on to the next unit treatment, A_1. Since A_1 is the last unit treatment in the sequence, its redrawn DAG is identical with the original DAG in Fig. 13.16a. Applying the backdoor criterion, we see that the total causal effect of A_1 on Y is identified by conditioning on $Z_1 = (A_0, L_1)$, which block all backdoor paths. As the proper causal paths that each unit treatment, A_0 and A_1, contributes to the joint causal effect are thus identified, the joint causal effect is identified.

If the sequential backdoor criterion is met, then the distribution of the potential outcomes after intervening to set $A_t = a_t$ for all t can be nonparametrically estimated by

$$P\left(y^{a_0,\ldots a_n}\right) = \sum\nolimits_{l_0,\ldots l_n} P\left(y|l_0,\ldots,l_n,a_0,\ldots,a_n\right) \times \prod\nolimits_{t=0}^{n} P\left(l_t|l_0,\ldots,l_{t-1},a_0,\ldots,a_{t-1}\right). \tag{13.7}$$

[23] A minimally sufficient set is a sufficient set with the smallest number of variables. There may be multiple minimally sufficient sets.

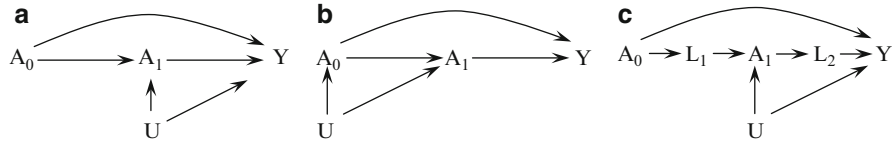

Fig. 13.17 All variables except U are observed. (**a**) The total causal effect of A_0 on Y is identified but the joint causal effect of A_0 and A_1 on Y is not. (**b**) The total causal effect of A_0 on Y is not identified but the joint causal effect of A_0 and A_1 on Y is identified. (**c**) The joint causal effect of A_0 and A_1 on Y is identified by the do-calculus even though the sequential backdoor criterion fails

From this distribution, one can compute all desired causal effects simply by setting the a_t to the desired values. Equation (13.7) is known as the *g-formula* in biostatistics (Robins 1986).

Unfortunately, the nonparametric estimator of Eq. (13.7) is often not practicable if there are many covariates or many treatment periods, or if the sample is small. As in Eq. (13.5), the analyst may be forced to insert parametric assumptions into the terms of the nonparametric estimator. Once more, things can go wrong at this stage. For example, conventional single-equation regression models, such as $Y = \alpha + \beta_0 A_0 + \beta_1 A_1 + \gamma_0 L_0 + \gamma_1 L_1 + e_Y$, often fail to provide unbiased estimates for the nonparametrically identified joint causal effects, especially if there are time-varying confounders, such as L_1 in Fig. 13.16. DAGs readily communicate the heart of the problem. Consider how one should handle L_1 in the analysis. On one hand, L_1 is a confounder of A_1, $A_1 \leftarrow L_1 \rightarrow Y$ and thus must be conditioned on. On the other hand, L_1 lies on a proper causal path from A_0 to Y, $A_0 \rightarrow L_1 \rightarrow Y$, such that conditioning on it would induce overcontrol bias. What is more, L_1 is also a collider on the noncausal path $A_0 \rightarrow L_1 \leftarrow U \rightarrow Y$, such that conditioning on it would open the noncausal path and induce endogenous selection bias. It is thus both necessary and forbidden to condition on L_1. In a conventional single-equation regression model, simultaneously conditioning and not conditioning on L_1 is impossible. Hence, conventional regression models may be biased for the joint causal effect of time-varying treatments even if the joint causal effect is in fact nonparametrically identified by the sequential backdoor criterion (Robins 1999).

For situations where (1) the joint causal effect is identified, (2) the nonparametric estimator is not practicable, and (3) the conventional regression methods fail, Robins and others have developed several more flexible parametric and semi-parametric estimators, such as structural nested models (Robins 1997) and marginal structural models with inverse probability of treatment weighting (Robins 1999). See Robins and Hernán (2009) for a comprehensive review. See Wodtke et al. (2011) for a sociological application to the joint effects of time-varying neighborhood conditions on education. See Sharkey and Elwert (2011) for an example of a formal sensitivity analysis for model violations (Robins 1999; Brumback et al. 2004).

A few further remarks on the sequential backdoor criterion can be helpful in practice. First, it is sometimes possible to assess the identification of the sequential backdoor criterion at a glance. Note, for example, that the sequential backdoor criterion is certainly met if there are no arrows from unobserved variables on the causal DAG into any unit treatment, as in Fig. 13.16. Such DAGs represent sequentially randomized experiments or observational studies in which all parents of all unit treatments are observed. Second, note that it may be possible to identify a joint causal effect even if the total causal effects of some unit treatments are not identified and vice versa. For example, in Fig. 13.17a, the total causal effect of A_0 on Y is identified even though the joint causal effect of A_0 and A_1 is not identified; and in Fig. 13.17b, the joint causal effect of A_0 and A_1 is identified even though the total causal effect of A_0 on Y is not identified. Finally, it is important to keep in mind that the sequential backdoor criterion is sufficient but not necessary for the identification of joint causal effects. For example, the joint causal effect of A_0 and A_1 on Y in Fig. 13.17c is not identified by the sequential backdoor criterion, but it is identifiable by the do-calculus.

Conclusion

The literature on DAGs is growing fast, both at the technical frontier and with respect to empirical applications. A central advantage of graphical causal models lies in combining rigor with transparency. DAGs enable applied researchers without years of mathematical training to assimilate and apply many previously inaccessible results. In part, practitioners seem to value DAGs as translation tools. Much as the ignorability criterion of the potential outcomes framework once helped methodologists make sense of regression-type approaches, so have the adjustment and backdoor criteria of the graphical framework helped analysts see ignorability in a new light (Pearl 2009: 341–43). Graphical approaches to causal inference are also drivers of methodological progress themselves. For example, Pearl's (1995) do-calculus gives nonparametric identification rules that are not only more general than ignorability but provably complete for the causal effects of interventions (Shpitser and Pearl 2006).

This chapter has introduced the central principles of working with DAGs and illustrated these principles with a number of methodological and substantive topics relevant for applied social science. Chief among the general principles are identification and model testing. DAGs, as visual representations of qualitative causal assumptions, encode the causal model needed for nonparametric identification analysis. Via d-separation, DAGs inform the analyst of all nonparametrically testable implications of their causal model. Using some sufficient graphical identification criteria (especially the adjustment criterion), this chapter then discussed several conceptual and applied topics in social research, including the causal nature of confounding and endogenous selection bias and the central identification problems of causal mediation and network analysis. Moving beyond simple covariate adjustment for single, fixed-time treatments, this chapter also showed how DAGs inform the nonparametric identification of joint causal effects from multiple, time-varying treatments.

One obvious challenge of working with DAGs is that the true causal DAG is often not known. This is a problem because identification always hinges on the validity of the causal model. If the DAG is incorrect, the identification conclusions drawn from it may be incorrect as well. It would be misguided, however, to blame DAGs for what ultimately are limitations of substantive scientific knowledge. The identification of causal effects requires causal assumptions regardless of how these assumptions are notated. DAGs seem to be especially well suited to draw attention to incomplete or implausible causal assumptions. This is a good thing. Assumptions do not disappear simply because they are hidden in a thicket of notation, and they cannot be corrected unless they are noticed and understood (Pearl 2009). One hopes that transparency might spur scientific progress.

A related problem is that the common-cause-inclusion requirement of causal DAGs quickly leads to unmanageably large DAGs in nonexperimental settings. Taking the shortcut of placing bi-headed arrows on pairs of variables that one suspects of being confounded usually leads to the disappearance of exclusion restrictions and hence to the realization that hardly any causal effect appears identifiable. This too, however, can hardly be blamed on the graphical framework per se, which merely reveals that poor theory (or strong theory coupled with poor data) rarely supports the identification of causal effects in observational studies.

Nevertheless, even unrealistically sparse DAGs can serve an important purpose in highlighting problems inherent in larger, more realistic DAGs built around them. For example, we do not need comprehensive theories of fertility, economic decision making in firms and families, and macroeconomic business cycles to understand why restricting the sample to employed women may bias estimates of the motherhood wage penalty. A simple DAG like that in Fig. 13.13 does the trick.

The frontier of technical research on DAGs today has moved on to topics beyond the scope of this chapter. Considerable work is being done on topics applied (e.g., the identification of various types of direct and indirect causal effects (Robins and Richardson 2011)), foundational (e.g., graphs for counterfactual variables (Shpitser and Pearl 2007)), and as yet arcane (e.g., auxiliary variable

identification (Chan and Kuroki 2010)). The foundational insights covered in this chapter, however, such as d-separation and the nonparametric identification of causal effects of interventions, are settled and have established a firm place in applied research.

Acknowledgments I thank Stephen Morgan, Judea Pearl, Tyler VanderWeele, Xiaolu Wang, Christopher Winship, and my students in Soc 952 at the University of Wisconsin for discussions and advice. Janet Clear and Laurie Silverberg provided editorial assistance. All errors are mine.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin, 2*(3), 47–53.

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Brito, C., & Pearl, J. (2002). Generalized instrumental variables. In A. Darwiche & N. Friedman (Eds.), *Uncertainty in artificial intelligence, proceedings of the eighteenth conference*(pp. 85–93). San Francisco: Morgan Kaufmann.

Brumback, B. A., Hernán, M. A., Haneuse, S. J. P. A., & Robins, J. M. (2004). Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Statistics in Medicine, 23*, 749–767.

Burns, B. D., & Wieth, M. (2004). The collider principle in causal reasoning: Why the Monty Hall Dilemma is so hard. *Journal of Experimental Psychology: General, 133*(3), 434–449.

Chan, H., & Kuroki, M. (2010). Using descendants as instrumental variables for the identification of direct causal effects in linear SEMs. In *Proceedings of the thirteenth international conference on Artificial Intelligence and Statistics* (AISTATS-10) (pp. 73–80), Sardinia, Italy.

Cole, S. R., & Hernán, M. A. (2002). Fallibility in estimating direct effects (with discussion). *International Journal of Epidemiology, 31*, 163–165.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic.

Elwert, F., & Christakis, N. A. (2006). Widowhood and race. *American Sociological Review, 71*(1), 16–41.

Elwert, F., & Christakis, N. A. (2008). Wives and ex-wives: A new test for homogamy bias in the widowhood effect. *Demography, 45*(4), 851–873.

Elwert, F., & Winship, C. (2010). Effect heterogeneity and bias in main-effects-only regression models. In R. Dechter, H. Geffner, & J. Y. Halpern (Eds.), *Heuristics, probability and causality: A tribute to Judea Pearl* (pp. 327–336). London: College Publications.

Elwert, F., & Winship, C. (forthcoming). Endogenous selection bias the dangers of conditioning on collider variables. *Annual Review of Sociology*.

Farr, W. (1858). Influence of marriage on the mortality of the French people. In G. W. Hastings (Ed.), *Transactions of the national association for the promotion of social science* (pp. 504–513). London: John W. Park & Son.

Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size. *American Educational Research Journal, 27*(3), 557–577.

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology, 97*(2), 214–223.

Fowler, J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *PNAS: Proceedings of the National Academy of Sciences, 107*(12), 5334–5338.

Galles, D., & Pearl, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science, 3*(1), 151–182.

Glymour, M. M., & Greenland, S. (2008). Causal diagrams. In K. J. Rothman, S. Greenland, & T. Lash (Eds.), *Modern epidemiology* (3rd ed., pp. 183–209). Philadelphia: Lippincott.

Greenland, S. (2003). Quantifying biases in causal models: Classical confounding versus collider-stratification bias. *Epidemiology, 14*, 300–306.

Greenland, S. (2010). Overthrowing the tyranny of null hypotheses hidden in causal diagrams. In R. Dechter, H. Geffner, & J. Y. Halpern (Eds.), *Heuristics, probability and causality: A tribute to Judea Pearl* (pp. 365–382). London: College Publications.

Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability and epidemiological confounding. *International Journal of Epidemiology, 15*, 413–419.

Greenland, S., Pearl, J., & Robins, J. M. (1999a). Causal diagrams for epidemiologic research. *Epidemiology, 10*, 37–48.

Greenland, S., Robins, J. M., & Pearl, J. (1999b). Confounding and collapsibility in causal inference. *Statistical Science, 14*, 29–46.

Gronau, R. (1974). Wage comparisons-a selectivity bias. *Journal of Political Economy, 82*, 1119–1144.

Heckman, J. J. (1974). Shadow prices, market wages and labor supply. *Econometrica, 42*(4), 679–694.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement, 5*, 475–492.

Hernán, M. A., Hernández-Diaz, S., Werler, M. M., Robins, J. M., & Mitchell, A. A. (2002). Causal knowledge as a prerequisite of confounding evaluation: An application to birth defects epidemiology. *American Journal of Epidemiology, 155*(2), 176–184.

Hernán, M. A., Hernández-Diaz, S., & Robins, J. M. (2004). A structural approach to section bias. *Epidemiology, 155*(2), 174–184.

Hernán, M. A., Clayton, D., & Keiding, N. (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology, 40*, 780–785.

Holland, P. W. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association, 81*, 945–970.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equation models. *Sociological Methodology, 18*, 449–484.

Kim, J.H., & Pearl, J. (1983). A computational model for combined causal and diagnostic reasoning in inference systems. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence* (pp. 190–193). Karlsruhe.

Kyono, T. (2010). *Commentator: A front-end user-interface module for graphical and structural equation modeling* (Tech. Rep. (R-364)). UCLA Cognitive Systems Laboratory.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge: Cambridge University Press.

Morgan, S. L., & Winship, C. (2012). Bringing context and variability back in to causal analysis. In H. Kincaid (Ed.), *Oxford handbook of the philosophy of the social sciences*. New York: Oxford University Press.

Neyman, J. ([1923] 1990). On the application of probability theory to agricultural experiments. Essay on principles, section 9, translated (with discussion). *Statistical Science, 5*(4), 465–480.

O'Malley, A. J., Elwert, F., Rosenquist, J. N., Zaslavsky, A. M., & Christakis, N. A. (2012). *Estimating peer effects in longitudinal dyadic data using instrumental variables* (Working Paper). Department of Health Care Policy, Harvard Medical School.

Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings, Cognitive Science Society* (pp. 329–334). Irvine: University of California.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufman.

Pearl, J. (1993). Comment: Graphical models, causality, and interventions. *Statistical Science, 8*(3), 266–269.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82*(4), 669–710.

Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods and Research, 27*(2), 226–284.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the seventeenth conference on Uncertainty in Artificial Intelligence* (pp. 411–420). San Francisco: Morgan Kaufmann.

Pearl, J. ([2000] 2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press.

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

Pearl, J. (2012a). The causal foundations of structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 68–91). New York: Guilford Press.

Pearl, J. (2012b). *Interpretable conditions for identifying direct and indirect effects* (Tech. Rep. (R-389)). UCLA Cognitive Systems Laboratory.

Pearl, J., & Robins, J. M. (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables. In P. Besnard & S. Hanks (Eds.), *Uncertainty in artificial intelligence 11* (pp. 444–453). San Francisco: Morgan Kaufmann.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period: Application to the health worker survivor effect. *Mathematical Modeling, 7*, 1393–1512.

Robins, J. M. (1989). The control of confounding by intermediate variables. *Statistics in Medicine, 8*, 679–701.

Robins, J. M. (1997). Causal inference from complex longitudinal data. In M. Berkane (Ed.), *Latent variable modeling and applications to causality* (Lecture notes in statistics 120, pp. 69–117). New York: Springer.

Robins, J. M. (1999). Association, causation, and marginal structural models. *Synthese, 121*, 151–179.

Robins, J. M. (2001). Data, design, and background knowledge in etiologic inference. *Epidemiology, 23*(3), 313–320.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology, 3*, 143–155.

Robins, J. M., & Hernán, M. A. (2009). Estimation of the causal effects of time-varying exposures. In G. Fitzmaurice et al. (Eds.), *Handbooks of modern statistical methods: Longitudinal data analysis* (pp. 553–599). Boca Raton: CRC Press.

Robins, J. M., & Richardson, T. (2011). Alternative graphical causal models and the identification of direct effects. In P. Shrout, K. Keyes, & K. Ornstein (Eds.), *Causality and psychopathology: Finding the determinants of disorders and their cures* (pp. 103–158). New York: Oxford University Press.

Robins, J. M., & Wasserman, L. (1999). On the impossibility of inferring causation from association without background knowledge. In C. N. Glymour & G. G. Cooper (Eds.), *Computation, causation, and discovery* (pp. 305–321). Cambridge: AAAI/MIT Press.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A, 147*(5), 656–666.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

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

Rubin, D. B. (1980). Comment on 'randomization analysis of experimental data in the fisher randomization test' by Basu. *Journal of the American Statistical Association, 75*, 591–593.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research, 40*, 211–239.

Sharkey, P., & Elwert, F. (2011). The legacy of disadvantage: Multigenerational neighborhood effects on cognitive ability. *The American Journal of Sociology, 116*(6), 1934–1981.

Shpitser, I., & Pearl, J. (2006). Identification of conditional interventional distributions. In R. Dechter & T. S. Richardson (Eds.), *Proceedings of the twenty-first national conference on Artificial Intelligence* (pp. 437–444). Menlo Park: AAAI Press.

Shpitser, I., & Pearl, J. (2007). What counterfactuals can be tested. In *Proceedings of the twenty-third conference on Uncertainty in Artificial Intelligence (UAI-07)* (pp. 352–359). Corvallis: AUAI Press.

Shpitser, I., VanderWeele, T. J., & Robins, J. M. (2010). On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the 26th conference on Uncertainty and Artificial Intelligence* (pp. 527–536). Corvallis: AUAI Press.

Shrier, I. (2009). Letter to the editor. *Statistics in Medicine, 27*, 2740–2741.

Sjölander, A. (2009). Letter to the editor: Propensity scores and M-structures. *Statistics in Medicine, 28*, 1416–1423.

Smith, H. L. (1990). Specification problems in experimental and nonexperimental social research. *Sociological Methodology, 20*, 59–91.

Sobel, M. E. (2008). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics, 33*(2), 230–251.

Spirtes, P., Glymour, C. N., & Schein, R. ([1993] 2001). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT Press.

Textor, J., Hardt, J., & Knüppel, S. (2011). Letter to the editor: DAGitty: A graphical tool for analyzing causal diagrams. *Epidemiology, 22*(5), 745.

VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology, 20*, 863–871.

VanderWeele, T. J. (2011). Sensitivity analysis for contagion effects in social networks. *Sociological Methods and Research, 40*, 240–255.

VanderWeele, T. J., & Robins, J. M. (2007). Four types of effect modification: A classification based on directed acyclic graphs. *Epidemiology, 18*(5), 561–568.

VanderWeele, T. J., & Robins, J. M. (2009). Minimal sufficient causation and directed acyclic graphs. *The Annals of Statistics, 37*, 1437–1465.

VanderWeele, T. J., & Shpitser, I. (2011). A new criterion for confounder selection. *Biometrics, 67*, 1406–1413.

Verma, T., & Pearl, J. (1988). Causal networks: Semantics and expressiveness. In *Proceedings of the fourth workshop on Uncertainty in Artificial Intelligence* (pp. 352–359). Minneapolis/Mountain View: AUAI Press.

Winship, C., & Harding, D. J. (2008). A mechanism-based approach to the identification of age-period-cohort models. *Sociological Methods and Research, 36*(3), 362–401.

Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *American Sociological Review, 76*, 713–736.

Wooldridge, J. (2005). Violating ignorability of treatment by controlling for too many factors. *Econometric Theory, 21*, 1026–1028.

Wooldridge, J. (2006). Acknowledgement of related prior work. *Econometric Theory, 22*, 1177–1178.