



# Wiley StatsRef: Statistics Reference Online



## Causal Diagrams

By *Sander Greenland and Judea Pearl*

**Keywords:** *bias, causal diagrams, causal inference, confounding, directed acyclic graphs, epidemiology, graphical models, path analysis, selection bias, standardization*

**Abstract:** From their inception, causal systems models (more commonly known as structural equations models) have been accompanied by graphical representations or path diagrams that provide compact summaries of qualitative assumptions made by the models. These diagrams can be reinterpreted as probability models, enabling use of graph theory in probabilistic inference, and allowing easy deduction of independence conditions implied by the assumptions. They can also be used as a formal tool for causal inference, such as predicting the effects of external interventions. Given that the diagram is correct, one can see whether the causal effects of interest (target effects or causal estimands) can be estimated from available data, or what additional observations are needed to validly estimate those effects. One can also see how to represent the effects as familiar standardized effect measures. This article gives an overview of the following: (i) components of causal graph theory; (ii) probability interpretations of graphical models; and (iii) methodological implications of the causal and probability structures encoded in the graph, such as sources of bias and the data needed for their control.

From their inception in the early twentieth century, causal systems models (more commonly known as *structural equations models*) have been accompanied by **graphical representations** or path diagrams that provide compact summaries of qualitative assumptions made by the models. Figure 1 provides a graph that would correspond to any system of five equations encoding these assumptions:

1. independence of  $A$  and  $B$
2. direct dependence of  $C$  on  $A$  and  $B$
3. direct dependence of  $E$  on  $A$  and  $C$
4. direct dependence of  $F$  on  $C$
5. direct dependence of  $D$  on  $B$ ,  $C$ , and  $E$ .

The interpretation of “direct dependence” was kept rather informal and usually conveyed by causal intuition, for example, that the entire influence of  $A$  on  $F$  is “mediated” by  $C$ .

University of California, Los Angeles, CA, USA

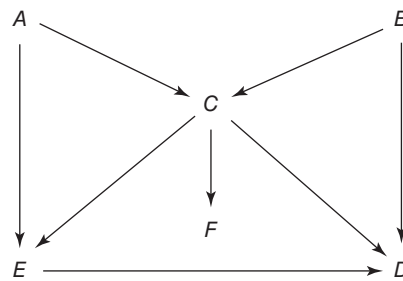
Update based on original article by Sander Greenland, Wiley StatsRef: Statistics Reference Online © 2014 John Wiley & Sons, Ltd.

Wiley StatsRef: Statistics Reference Online, © 2014–2017 John Wiley & Sons, Ltd.

This article is © 2017 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat03732.pub2





**Figure 1.** Example of causal diagram.

By the 1980s, it was recognized that these diagrams could be reinterpreted formally as probability models, enabling use of graph theory in probabilistic inference, and allowing easy deduction of independence conditions implied by the assumptions<sup>[1]</sup>. By the 1990s, it was further recognized that these diagrams could also be used as a formal tool for causal inference<sup>[2–5]</sup> and for illustrating sources of bias and their remedy in epidemiological research<sup>[6–15]</sup>. Given that the graph is correct, one can analyze whether the causal effects of interest (target effects or causal estimands) can be estimated from available data, or what additional observations are needed to validly estimate those effects. One can also see from the graph which variables should and should not be conditioned on (controlled) in order to estimate a given effect.

This article gives an overview of the following: (i) components of causal graph theory, (ii) probability interpretations of graphical models, and (iii) methodological implications of the causal and probability structures encoded in the graph, such as sources of bias and the data needed for their control. See **Causality/Causation** for a discussion of definitions of causation and statistical models for causal inference.

## 1 Basics of Graph Theory

As a befitting and well-developed mathematical topic, graph theory has an extensive terminology that once mastered provides access to a number of elegant results that may be used to model any system of relations. The term *dependence* in a graph, usually represented by connectivity, may refer to mathematical, causal, or statistical dependencies. The connectives joining variables in the graph are called *arcs*, *edge*, or *links*, and the variables are also called *nodes* or *vertices*. Two variables connected by an arc are *adjacent* or *neighbors* and arcs that meet at a variable are also adjacent. If the arc is an arrow, the tail (starting) variable is the *parent* and the head (ending) variable is the *child*. In causal diagrams, an arrow represents a “direct effect” of the parent on the child, although this effect is direct only relative to a certain level of abstraction, in that the graph omits any variables that might mediate the effect represented by the arrow.

A variable that has no parent (such as *A* and *B* in Figure 1) is *exogenous* or *external*, or a *root* or *source* node, and is determined only by forces outside the graph; otherwise, it is *endogenous* or *internal*. A variable with no children (such as *D* in Figure 1) is a *sink* or *terminal node*. The set of all parents of a variable *X* (all variables at the tail of an arrow pointing into *X*) is denoted  $\text{pa}[X]$ ; in Figure 1,  $\text{pa}[D] = \{B, C, E\}$ .

A *path* or *chain* is a sequence of adjacent arcs. A *directed path* is a path traced out entirely along arrows tail-to-head. If there is a directed path from *X* to *Y*, *X* is an *ancestor* of *Y* and *Y* is a *descendant* of *X*. In causal diagrams, directed paths represent causal pathways from the starting variable to the ending variable; a variable is thus often called a *cause of its descendants* and an *effect of its ancestors*. In a *directed* graph, the



only arcs are arrows, and in an *acyclic* graph, there is no feedback loop (no directed path from a variable back to itself). Therefore, a directed acyclic graph (DAG) is a graph with only arrows for edges and no feedback loops (i.e., no variable is its own ancestor or its own descendant). A causal DAG represents a complete causal structure in that all sources of dependence are explained by causal links; in particular, all common (shared) causes of variables in the graph are also in the graph.

A variable *intercepts* a path if it is in the path (but not at the ends); similarly, a set of variables  $S$  intercepts a path if it contains any variable intercepting the path. Variables that intercept directed paths are *intermediates* or *mediators* on the pathway. A variable is a *collider* on the path if the path enters and leaves the variable via arrowheads (a term suggested by the collision of causal forces at the variable). Note that being a collider is relative to a path; for example, in Figure 1,  $C$  is a collider on the path  $A \rightarrow C \leftarrow B \rightarrow D$  and a noncollider on the path  $A \rightarrow C \rightarrow D$ . Nonetheless, it is common to refer to a variable as a collider if it is a collider along any path (i.e., if it has more than one parent). A path is *open* or *unblocked* at noncolliders and *closed* or *blocked* at colliders; hence, a path with no collider (such as  $E \leftarrow C \leftarrow B \rightarrow D$ ) is *open* or *active*, while a path with a collider (such as  $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$ ) is *closed* or *inactive*.

Some of the most important constraints imposed by a graphical model correspond to independencies arising from the absence of open paths; for example, the absence of an open path from  $A$  to  $B$  in Figure 1 constrains  $A$  and  $B$  to be marginally independent (i.e., independent if no stratification is done). Nonetheless, the converse does not hold; that is, the presence of an open path allows but does not imply dependency. Independence may arise through the cancellation of dependencies; consequently, even adjacent variables may be marginally independent; for example, in Figure 1,  $A$  and  $E$  could be marginally independent if the dependencies through paths  $A \rightarrow E$  and  $A \rightarrow C \rightarrow E$  cancelled each other. The assumption of faithfulness, discussed in the following paragraph, excludes such possibilities.

Some authors use a bidirectional arc (two-headed arrow,  $\leftrightarrow$ ) to represent the assumption that two variables share ancestors that are not shown in the graph;  $A \leftrightarrow B$  then means that there is an unspecified variable  $U$  with directed paths to both  $A$  and  $B$  (e.g.,  $A \leftarrow U \rightarrow B$ ). Certain authors use  $A \leftrightarrow B$  less specifically, however, in the way  $A - B$  is used here to designate the presence of nondirected open paths between  $A$  and  $B$  that are not intercepted by any variable in the graph.

### 1.1 Control: Manipulation Versus Conditioning

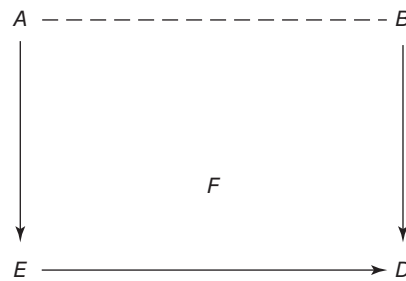
---

The word *control* is used throughout science but with a variety of meanings that are important to distinguish. In experimental research, to control a variable  $C$  usually means to manipulate or set its value. In observational studies, however, to control  $C$  (or more precisely, to control for  $C$ ) more often means to condition on  $C$ , usually by stratifying on  $C$  or by entering  $C$  in a regression model. The two processes are very different physically and have very different representations and implications.

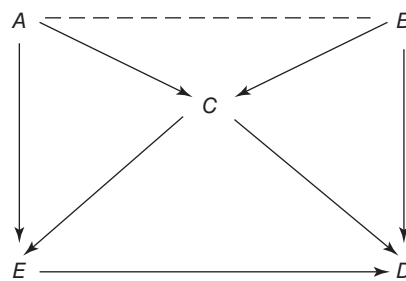
If a variable  $X$  is influenced by a researcher, the DAG would need an ancestor  $R$  of  $X$  to represent this influence. In the classical experimental case in which the researcher alone determines  $X$ , the variables  $R$  and  $X$  would be identical. In human trials, however,  $R$  more often represents just an *intention* to treat ( $R$  is only the *assigned* level of  $X$ ), leaving  $X$  to be influenced by other factors that affect compliance with the assigned treatment  $R$ . In either case,  $R$  might be affected by other variables in the graph. For example, if the researcher uses age to determine assignments (an age-biased allocation), age would be a parent of  $R$ . Ordinarily, however,  $R$  would be exogenous as when  $R$  represents a randomized allocation.

In contrast, by definition in an observational study, there is no such variable  $R$  representing the researcher's influence on  $X$ , and conditioning is substituted for experimental control. Conditioning on a variable  $C$  in a DAG can be represented by creating a new graph from the original graph to represent constraints on relations within levels (strata) of  $C$  implied by the constraints imposed by the original





**Figure 2.** Graph derived from Figure 1 after conditioning on  $C$ .



**Figure 3.** Graph derived from Figure 1 after conditioning on  $F$ .

graph. This conditional graph can be found by the following sequence of operations<sup>[6]</sup>, sometimes called *graphical moralization*.

1. If  $C$  is a collider, join (marry) all pairs of parents of  $C$  by undirected arcs (here, a dashed line will be used).
2. Similarly, if  $A$  is an ancestor of  $C$  and a collider, join all pairs of parents of  $A$  by undirected arcs.
3. Erase  $C$  and all arcs connecting  $C$  to other variables.

Figure 2 shows the graph derived from conditioning on  $C$  in Figure 1: the parents  $A$  and  $B$  of  $C$  are joined by an undirected arc, while  $C$  and all its arcs are gone. Figure 3 shows that the result of conditioning on  $F$ :  $C$  is an ancestral collider of  $F$  and so, again its parents  $A$  and  $B$  are joined, but only  $F$  and its single arc are erased. Note that, because of the undirected arcs, neither figure is a DAG.

Operations 1 and 2 reflect that if  $C$  depends on  $A$  and  $B$  through distinct pathways, the marginal dependence of  $A$  on  $B$  will not equal the dependence of  $A$  on  $B$  stratified on  $C$  (apart from special cases). To illustrate, suppose that  $A$  and  $B$  are binary indicators (i.e., equal to 1 or 0), marginally independent, and  $C = A + B$ . Then, among persons with  $C = 1$ , some will have  $A = 1, B = 0$  and some will have  $A = 0, B = 1$  (because other combinations produce  $C \neq 1$ ). Thus, when  $C = 1$ ,  $A$  and  $B$  will exhibit perfect negative dependence:  $A = 1 - B$  for all persons with  $C = 1$ .

Conditioning on a variable  $C$  closes open paths that pass through  $C$ . Conversely, conditioning on  $C$  opens paths that were blocked only at  $C$  or at an ancestral collider  $A$ . In particular, conditioning on a variable may open a path even if it is not on the path, as with  $F$  in Figures 1 and 3. Thus, the consequences of conditioning on  $C$  are often illustrated in the original (unconditional) graph by drawing a circle around  $C$  to denote the conditioning, then defining a path blocked by the circled  $C$  if  $C$  is a noncollider on the path, and by “marrying the parents” of  $C$  and its collider ancestors (as in 1 and 2 above) to show the paths



opened by conditioning. Thus, if we circle  $C$  in Figure 1, it will block the  $E$ - $D$  paths  $E \leftarrow C \leftarrow B \rightarrow D$  and  $E \leftarrow A \rightarrow C \rightarrow D$  but unblock the path  $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$  via the new arc between  $A$  and  $B$ . If we were to circle  $F$  but not  $C$ , no open path would be blocked, but  $A$  and  $B$  would again be joined by a dashed arc as in Figure 3.

A path is closed after conditioning on a set of variables  $S$  if  $S$  contains a noncollider along the path or if the conditioning leaves the path closed at a collider; in either case  $S$  is said to block the path. Thus, conditioning on  $S$  closes an open path if and only if  $S$  intercepts path and opens a closed path if  $S$  contains no noncolliders on the path and every collider on the path is either in  $S$  or has a descendant in  $S$ . In Figure 1, the closed path  $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$  will remain closed after conditioning on  $S$  if  $S$  contains  $A$  or  $B$  or if  $S$  does not contain  $C$  but will be opened if  $S$  contains only  $C$ ,  $F$ , or both.

Two variables (or sets of variables) in the graph are *d-separated* (or separated in the DAG) by a set  $S$  if, after conditioning on  $S$ , there is no open path between them. Thus, in Figure 1,  $\{A, C\}$  separates  $E$  from  $B$ , but  $\{C\}$  does not because conditioning on  $C$  alone results in Figure 2, in which  $E$  and  $B$  are connected via the open path  $A$ . In a DAG,  $\text{pa}[X]$  *d-separates*  $X$  from every variable that is not affected by  $X$  (i.e., not a descendant of  $X$ ). This feature of DAGs is sometimes called the *Markov condition*, expressed by saying the parents of a variable “screen off” the variable from everything but its effects. Thus, in Figure 1,  $\text{pa}[E] = \{A, C\}$ , which separates  $E$  from  $B$  but not from  $D$ .

## 1.2 Bias and Confounding

There is considerable variation in the literature in the usage of terms such as “**Bias**,” “confounding,” and related concepts, which refer to dependencies that reflect more than just the effect under study. To capture these notions in a causal graph, we say that an open path between  $X$  and  $Y$  is a *biasing path* if it is not a directed path. The association of  $X$  with  $Y$  is then *unbiased* for the effect of  $X$  on  $Y$  if the only open paths from  $X$  to  $Y$  are the directed paths. Similarly, the dependence of  $Y$  on  $X$  is *unbiased given  $S$*  if, after conditioning on  $S$ , the open paths between  $X$  and  $Y$  are exactly (only and all) the directed paths in the starting graph. In such a case, we say  $S$  is *sufficient* to block bias in the  $X$ - $Y$  dependence and is *minimally sufficient* if no proper subset of  $S$  is sufficient.

Informally, confounding is a source of bias arising from causes of  $Y$  that are associated with but not affected by  $X$  (see **Confounding**). Thus, we say an open nondirected path from  $X$  to  $Y$  is a *confounding path* if it ends with an arrow into  $Y$ . Variables that intercept confounding paths between  $X$  and  $Y$  are *confounders*. If a confounding path is present, we say *confounding* is present and that the dependence of  $Y$  on  $X$  is *confounded*. If no confounding path is present, we say the dependence is *unconfounded*, in which case the only open paths from  $X$  to  $Y$  through a parent of  $Y$  are directed paths. Similarly, the dependence of  $Y$  on  $X$  is *unconfounded given  $S$*  if, after conditioning on  $S$ , the only open paths between  $X$  and  $Y$  through a parent of  $Y$  are directed paths.

An unconfounded dependency may still be biased owing to nondirected open paths that do not end in an arrow into  $Y$ . These paths can be created when one conditions on a descendant of both  $X$  and  $Y$ . The resulting bias is called *Berksonian bias*, after its discoverer Joseph Berkson<sup>[16]</sup>. Most epidemiologists call this type of bias as “selection bias”<sup>[16]</sup>. Nonetheless, some writers (especially in econometrics) use “selection bias” to refer to confounding, while others call any bias created by conditioning as “selection bias”<sup>[14]</sup>.

Consider a set of variables  $S$  that contains no effect (descendant) of  $X$  or  $Y$ .  $S$  is *sufficient* to block confounding if the dependence of  $Y$  on  $X$  is unconfounded given  $S$ . “No confounding” thus corresponds to sufficiency of the empty set. A sufficient  $S$  is called *minimally sufficient* to block confounding if no proper subset of  $S$  is sufficient. The initial exclusion from  $S$  of descendants of  $X$  or  $Y$  in these definitions arises first because conditioning on  $X$  descendants can easily block directed (causal) paths that are part of the



effect of interest and, second, because conditioning on  $X$  or  $Y$  descendants can unblock paths that are not part of the  $X$ – $Y$  effect and thus create new bias.

A *back-door path* from  $X$  to  $Y$  is a path that begins with a parent of  $X$  (i.e., leaves  $X$  from a “back door”) and ends at  $Y$ . A set  $S$  then satisfies the *back-door criterion* with respect to  $X$  and  $Y$  if (a)  $S$  contains no descendant of  $X$  and (b) there are no open back-door paths from  $X$  to  $Y$  after conditioning on  $S$ . In a DAG, the following simplifications occur.

1. All biasing paths are back-door paths.
2. The dependence of  $Y$  on  $X$  is unbiased whenever there are no open back-door paths from  $X$  to  $Y$ .
3. If  $X$  is exogenous, the dependence of any  $Y$  on  $X$  is unbiased.
4. All confounders are ancestors of either  $X$  or  $Y$ .
5. A back-door path is open if and only if it contains a common ancestor of  $X$  and  $Y$ .
6. If  $S$  satisfies the back-door criterion, then  $S$  is sufficient to block  $X$ – $Y$  confounding.

These conditions do not extend to non-DAGs like Figure 2. In addition, although  $\text{pa}[X]$  always satisfies the back-door criterion and hence is sufficient in a DAG, it may be far from minimal sufficient. For example, there is no confounding and hence no need for conditioning whenever  $X$  separates  $\text{pa}[X]$  from  $Y$  (i.e., whenever the only open paths from  $\text{pa}[X]$  to  $Y$  are through  $X$ ).

As a final caution, we note that the biases dealt with by the abovementioned concepts are only confounding and selection biases. Biases due to measurement error and model-form misspecification require further structure to describe them<sup>[14,17]</sup>.

## 2 Statistical Interpretations and Applications

A joint probability distribution for the variables in a graph is *compatible* with the graph if two sets of variables are independent given  $S$  whenever  $S$  separates them. For such distributions, two sets of variables will be statistically unassociated if there is no open path between them. Many special results follow for distributions compatible with a DAG. For example, if in a DAG,  $X$  is not an ancestor of any variable in a set  $T$ , then  $T$  and  $X$  will be independent given  $\text{pa}[X]$ . A distribution compatible with a DAG can thus be reduced to a product of factors  $\text{Pr}(x|\text{pa}[X])$ , with one factor for each variable  $X$  in the DAG; this is sometimes called the *Markov factorization* for the DAG. When  $X$  is a treatment, this condition implies that the probability of treatment is fully determined by the parents of  $X$ ,  $\text{pa}[X]$ .

Suppose that now we are interested in the effect of  $X$  on  $Y$  in a DAG, and we assume a probability model compatible with the DAG. Then, given a sufficient set  $S$ , the only source of association between  $X$  and  $Y$  within strata of  $S$  will be the directed paths from  $X$  to  $Y$ . Hence, the *net effect* of  $X = x_1$  versus  $X = x_0$  on  $Y$  when  $S = s$  is defined as  $\text{Pr}(y|x_1, s) - \text{Pr}(y|x_0, s)$ , the difference in risks of  $Y = y$  at  $X = x_1$  and  $X = x_0$ . Alternatively, one may use another effect measure such as the risk ratio  $\text{Pr}(y|x_1, s)/\text{Pr}(y|x_0, s)$ . A *standardized effect* is a difference or ratio of weighted averages of these stratum-specific  $\text{Pr}(y|x, s)$  over  $S$ , using a common weighting distribution<sup>[18]</sup>. The latter definition can be generalized to include intermediate variables in  $S$  by allowing the weighting distribution to causally depend on  $X$ . Furthermore, given a set  $Z$  of intermediates along all directed paths from  $X$  to  $Y$  with  $X$ – $Z$  and  $Z$ – $Y$  unbiased, one can produce formulas for the  $X$ – $Y$  effect as a function of the  $X$ – $Z$  and  $Z$ – $Y$  effects (“front-door adjustment”<sup>[2,4]</sup>).

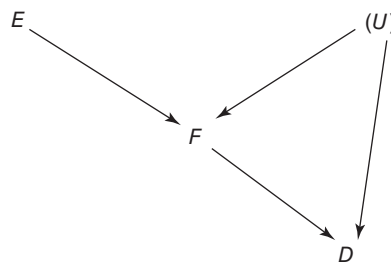
The abovementioned form of standardized effect is identical to the forms derived under other types of causal models, such as potential outcome models (see **Causation**). When  $S$  is sufficient, some authors identify the  $\text{Pr}(y|x, s)$  with the distribution of potential outcomes given  $S$ <sup>[4]</sup>. In that case, sufficiency of a set  $S$  implies that the potential outcome distribution equals  $\sum_s \text{Pr}(y|x, s)\text{Pr}(s)$ , the risk of  $Y = y$  given  $X = x$  standardized to the  $S$  distribution.

## 2.1 Identification of Effects and Biases

In order to check the sufficiency and identify minimally sufficient sets of variables given a graph of the causal structure, one needs to see only whether the open paths from  $X$  to  $Y$  after conditioning are exactly the directed paths from  $X$  to  $Y$  in the starting graph. Mental effort may then be shifted to evaluating the reasonableness of the causal independencies encoded by the graph, some of which are reflected in conditional independence relations. This property of graphical analysis facilitates the articulation of necessary background knowledge for estimating effects and eases teaching of algebraically difficult identification conditions.

As an example, spurious sample associations may arise if each variable affects selection into the study, even if those selection effects are independent. This phenomenon is a special case of the collider stratification effect illustrated earlier. Its presence is easily seen by starting with a DAG that includes a selection indicator  $F = 1$  for those selected, or  $z$  otherwise, as well as the study variables, and then noting that we are always forced to examine associations within the  $F = 1$  stratum (i.e., by definition, our observations stratify on selection). Thus, if selection ( $F$ ) is affected by multiple causal pathways, we should expect selection to create or alter associations among the variables. Graphical tools for analysis of selection bias are further described in Bareinboim *et al.*<sup>[19]</sup>.

Figure 4 displays a situation common in randomized trials, in which the net effect of  $E$  on  $D$  is unconfounded, despite the presence of an unmeasured cause  $U$  of  $D$ . Unfortunately, a common practice in health and social sciences is to stratify on (or otherwise adjust for) an intermediate variable  $F$  between a cause  $E$  and effect  $D$  and then claim that the estimated ( $F$ -residual) association represents that portion of the effect of  $E$  on  $D$  not mediated through  $F$ . In Figure 4, this would be a claim that, upon stratifying on  $F$ , the  $E$ - $D$  association represents the direct effect of  $E$  on  $D$ . Figure 5, however, shows the graph conditional on  $F$ , in which we see that there is now an open path from  $E$  to  $D$  through  $U$ , and hence, the residual  $E$ - $D$  association is confounded for the direct effect of  $E$  on  $D$ .



**Figure 4.** Graph with randomized  $E$  and confounded intermediate  $F$ .



**Figure 5.** Graph derived from Figure 4 after conditioning on  $F$ .

The  $E$ – $D$  confounding by  $U$  in Figure 5 can be seen as arising from the confounding of the  $F$ – $D$  association by  $U$  in Figure 4. In a similar manner, conditioning on  $C$  in Figure 1 opens the confounding path through  $A$  and  $B$  in Figure 2; this path can be seen as arising from the confounding of the  $C$ – $E$  association by  $A$  and the  $C$ – $D$  association by  $B$  in Figure 1. In both examples, further stratification on either  $A$  or  $B$  blocks the created path and thus removes the new confounding.

Bias from conditioning on a collider or its descendant has been called collider stratification bias or simply *collider bias*<sup>[12,15]</sup>. Starting from a DAG, there are two distinct forms of this bias: confounding induced in the conditional graph (Figures 2, 3, and 5) and Berksonian bias from conditioning on an effect of  $X$  and  $Y$ . Both biases can, in principle, be removed by further conditioning on variables along the biasing paths from  $X$  to  $Y$  in the conditional graph. Nonetheless, the starting DAG will always display ancestors of  $X$  or  $Y$  that, if known, could be used to remove confounding; in contrast, no variable need appear or even exist that could be used to remove Berksonian bias.

Figure 4 also provides a schematic for estimating the  $F$ – $D$  effect, as in randomized trials in which  $E$  represents assignment to or encouragement toward treatment  $F$ . One can put bounds on confounding of the  $F$ – $D$  association and with additional assumptions remove it entirely through formulas that treat  $E$  as an *instrumental variable* (a variable associated with  $F$  such that every open path to  $D$  includes an arrow pointing into  $F$ )<sup>[4,7,15]</sup>.

## 2.2 Questions of Discovery

While deriving statistical implications of graphical models is uncontroversial, algorithms that claim to discover causal (graphical) structures from observational data have been subject to strong criticism<sup>[20,21]</sup>. A key assumption in certain “discovery” algorithms is a converse of compatibility called *faithfulness*<sup>[5]</sup>. A compatible distribution is *faithful* to the graph (or *stable*) if for all  $X$ ,  $Y$ , and  $S$ ,  $X$  and  $Y$  are independent given  $S$  only when  $S$  separates  $X$  and  $Y$  (i.e., the distribution entails no independencies other than those implied by graphical separation). Faithfulness implies that minimal sufficient sets in the graph will also be minimal for consistent estimation of effects. Nonetheless, there are real examples of near cancellation (e.g., when confounding obscures a real effect), which make faithfulness questionable as a routine assumption. Fortunately, faithfulness is not needed for the uses of the graphical models discussed in this article.

Whether or not one assumes faithfulness, the generality of graphical models is purchased with limitations on their informativeness. Causal diagrams show whether the effects can be estimated (identified) from given information<sup>[2–7]</sup> and can be extended to indicate effect direction when that is monotone<sup>[22]</sup>. They can also be used to derive the qualitative impact of adjustments on association and effect measures<sup>[18]</sup>.

Nonetheless, the nonparametric nature of purely graphical models implies that parametric concepts such as linearity, additivity, and effect homogeneity cannot be displayed by the basic graphical theory reviewed earlier and so need to be specified separately, although qualitative properties can be represented<sup>[18,22,23]</sup>. Similarly, the graphs may imply that several distinct conditionings are minimal sufficient (e.g., both  $\{A, C\}$  and  $\{B, C\}$  are sufficient for the  $ED$  effect in Figure 1) but offer no further guidance on which to use. Open paths may suggest the presence of an association but that association may be negligible even if nonzero or may be unfaithfully constrained to zero by balanced matching<sup>[24]</sup>. For example, bounds on the size of direct effects imply more severe bounds on the size of effects mediated in multiple steps (indirect effects), with the bounds becoming more severe with each step. Consequently, there is often good reason to expect certain phenomena (such as the conditional  $E$ – $D$  confounding shown in Figures 2, 3, and 5) to be small in epidemiological examples. Finally, when (as is usually the case) the graphs represent properties of distributions rather than samples, they will not capture statistical artifacts such as random confounding and sparse-data bias<sup>[24]</sup>.





### 3 Other Reviews

Full technical details of causal diagrams and their relation to causal modeling and inference can be found in the books by Pearl<sup>[4]</sup> and Spirtes *et al.*<sup>[5]</sup>; Pearl<sup>[25]</sup> and Bareinboim and Pearl<sup>[26]</sup> further discuss relations between graphical and potential outcome models. Less technical reviews geared toward research scientists include Greenland *et al.*<sup>[6]</sup>, Greenland and Brumback<sup>[9]</sup>, Glymour and Greenland<sup>[15]</sup>, Pearl *et al.*<sup>[13]</sup>, and Hernán and Robins<sup>[27]</sup>.

### Related Articles

**Causation; Compliance with Treatment Allocation.**

### References

- [1] Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publishers, San Mateo.
- [2] Pearl, J. (1995) Causal diagrams for empirical research (with discussion). *Biometrika*, **82**, 669–710.
- [3] Pearl, J. and Robins, J.M. (1995) Probabilistic evaluation of sequential plans from causal models with hidden variables, in *Uncertainty in Artificial Intelligence* (eds P. Besnard and S. Hanks), Morgan Kaufmann Publishers, San Francisco, pp. 444–453.
- [4] Pearl, J. (2009) *Causality*, 2nd edn, Cambridge University Press, New York.
- [5] Spirtes, P., Glymour, C., and Scheines, R. (2001) *Causation, Prediction, and Search*, 2nd edn, MIT Press, Cambridge.
- [6] Greenland, S., Pearl, J., and Robins, J.M. (1999) Causal diagrams for epidemiologic research. *Epidemiology*, **10**, 37–48.
- [7] Greenland, S. (2000) An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.*, **29**, 722–729. (Erratum: 2000, **29**, 1102)
- [8] Robins, J.M. (2001) Data, design, and background knowledge in etiologic inference. *Epidemiology*, **12**, 313–320.
- [9] Greenland, S. and Brumback, B.A. (2002) An overview of relations among causal modelling methods. *Int. J. Epidemiol.*, **31**, 1030–1037.
- [10] Cole, S. and Hernán, M.A. (2002) Fallibility in estimating direct effects. *Int. J. Epidemiol.*, **31**, 163–165.
- [11] Hernán, M.A., Hernandez-Diaz, S., Werler, M.M., and Mitchell, A.A. (2002) Causal knowledge as a prerequisite for confounding evaluation. *Am. J. Epidemiol.*, **155**, 176–184.
- [12] Greenland, S. (2003) Quantifying biases in causal models: classical confounding versus collider-stratification bias. *Epidemiology*, **14**, 300–306.
- [13] Pearl, J., Glymour, M.M., and Jewell, N.P. (2016) *Causal Inference in Statistics: A Primer*, Wiley, Chapter 2, New York, pp. 35–52.
- [14] Hernán, M.A., Hernandez-Diaz, S., and Robins, J.M. (2004) A structural approach to selection bias. *Epidemiology*, **15**, 615–625.
- [15] Glymour, M.M. and Greenland, S. (2008) Causal diagrams, in *Modern Epidemiology*, 3rd edn (eds K.J. Rothman, S. Greenland, and T.L. Lash), Lippincott Williams & Wilkins, Philadelphia, Chapter 12.
- [16] Rothman, K.J., Greenland, S., and Lash, T.L. (2008) *Modern Epidemiology*, 3rd edn, Lippincott Williams & Wilkins, Philadelphia.
- [17] Hernán, M.A. and Cole, S. (2009) Causal diagrams and measurement bias. *Am. J. Epidemiol.*, **170**, 959–962.
- [18] Greenland, S. and Pearl, J. (2011) Adjustments and their consequences – collapsibility analysis using graphical models. *Int. Stat. Rev.*, **79**, 401–426.
- [19] Bareinboim, E., Tian J. and Pearl, J. (2014) *Recovering from Selection Bias in Causal and Statistical Inference*. Proceedings of the 28th AAAI Conference on Artificial Intelligence. AAAI Press, Palo Alto, CA.
- [20] Freedman, D.A. and Humphreys, P. (1999) Are there algorithms that discover causal structure? *Synthese*, **121**, 29–54.
- [21] Robins, J.M. and Wasserman, L. (1999) On the impossibility of inferring causation from association without background knowledge, in *Computation, Causation, and Discovery* (eds C. Glymour and G. Cooper), AAAI Press/The MIT Press, Menlo Park, pp. 305–321.
- [22] VanderWeele, T.J. and Robins, J.M. (2010) Signed directed acyclic graphs for causal inference. *J. R. Stat. Soc. Ser. B*, **72**, 111–127.



- [23] VanderWeele, T.J. and Robins, J.M. (2007) Four types of effect modification: a classification based on directed acyclic graphs. *Epidemiology*, **18**, 561–568.
- [24] Greenland, S. and Mansournia, M.A. (2016) Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. *Eur. J. Epidemiol.*, **30**, 1101–1110.
- [25] Pearl, J. (2014) The deductive approach to causal inference. *J. Causal Inference*, **2** (2), 115–129.
- [26] Bareinboim, E. and Pearl, J. (2016) Causal inference and the data-fusion problem. *Proc. Natl. Acad. Sci.*, **113** (27), 7345–7352.
- [27] Hernán, M.A. and Robins, J.M. (2017) *Causal Inference*, Chapman & Hall, New York.