Article prepared for the *Encyclopedia of Epidemiology*.

TECHNICAL REPORT
R-332
June 2006

CAUSAL DIAGRAMS

Sander Greenland
Epidemiology Department Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
lesdomes@ucla.edu

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

From their inception in the early 20th century, causal systems models (more commonly known as structural-equations models) were accompanied by graphical representations or path diagrams that provided compact summaries of qualitative assumptions made by the models. Fig. 1 provides a graph that would correspond to any system of 5 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 and,
5. direct dependence of D on B, C, and E

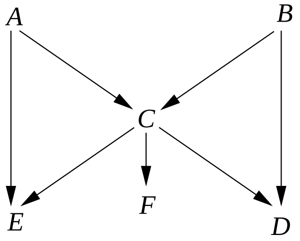

Figure 1:

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.

By the 1980s it was recognized that these diagrams could be reinterpreted formally as probability models, which opened the visual power of graph theory for use in probabilistic inference and allowed easy deduction of other independence conditions implied by the assumptions. By the 1990's it was further recognized that these diagrams could also be used as a formal tool for causal inference, such as predicting the effects of external interventions. Given that the graph 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.

The present article gives an overview of: (1) components of causal graph theory; (2) probability interpretations of graphical models; and (3) the methodologic implications of the causal and probability structures encoded in the graph. See CAUSATION AND CAUSAL INFERENCE for discussion of definitions of causation and statistical models for causal inference.

Basics of Graph Theory

As befitting a well developed mathematical topic, graph theory has an extensive terminology that, once mastered, provides access to a number of elegant results which 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.

A variable that has no parent (such as A and B in Fig. 1) is *exogenous* or *external*, or a *root* or *source* node, and is determined only by forces outside of the graph; otherwise it is *endogenous* or *internal*. A variable with no children (such as D in Fig. 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 $pa[X]$; in Fig. 1, $pa[D] = \{B, C, E\}$.

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 an *acyclic* is a graph in which there are no feedback loops (directed paths from a variable back to itself). Therefore, a directed acyclic graph or 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 DAG represents a complete causal structure, in that all sources of dependence are explained by causal links.

A variable *intercepts* or *mediates* 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* 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 Fig. 1, C is a collider on the path $A \to C \leftarrow B \to D$ and a noncollider on the path $A \to C \to D$. 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 (like $E \leftarrow C \leftarrow B \to D$) is *open* or *active*, while a path with a collider (like $E \leftarrow A \to C \leftarrow B \to D$) is *closed* or *inactive*.

Two variables (or sets of variables) in the graph are *d-separated* (or just separated) if there is no open path between them. Some of the most important constraints imposed by a graphical model correspond to independencies arising from separation; e.g., absence of an open path from A to B in Fig. 1 constrains A and B to be marginally independent (i.e., independent if no stratification is done). Nonetheless, the converse does not hold; i.e., presence of an open path allows but does not imply dependency. Independence may arise through cancellation of dependencies; as a consequence even adjacent variables may be marginally independent; e.g., in Fig. 1, A and E could be marginally independent if the dependencies through paths $A \to E$ and $A \to C \to E$ cancelled each other. The assumption of faithfulness, discussed below, is designed to exclude 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., $A \leftarrow U \to B$).

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 more often means to condition on C, usually by stratifying on C or to entering it 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, R and X would be identical. In human trials, however, R more often represents just an *intention* to treat (with 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 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 graph. This conditional graph can be found by following sequence of operations:

1. If C is a collider, join ("marry") all pairs of parents of C by undirected arcs; here dashed lines without arrowheads will be used (some authors use solid lines without arrowheads).

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.

Fig. 2 shows the graph derived from conditioning on C in Fig. 1: The parents A and B of C are joined by an undirected arc, while C and all its arcs are gone. Fig. 3 shows 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.

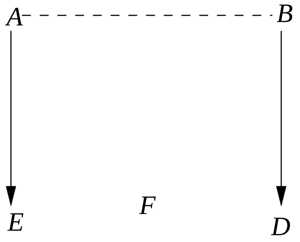

Figure 2:

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 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 reverses the status of C on paths that pass through it: Paths that were open at C are closed by conditioning on C, while paths that were closed at C become open at C (although they may remain closed elsewhere). Similarly, conditioning on a descendant of C partially reverses the status of C: Typically, paths that were open at C remain open, but with attenuated association across the path; while paths that were closed at C become open at C, although not as open as when conditioning on C itself. In other words, conditioning on a variable tends to partially reverse the status of ancestors on paths passing through the ancestors. In particular, conditioning on a variable may open a path even if it is not on the path, as with F in Fig. 1 and 3.

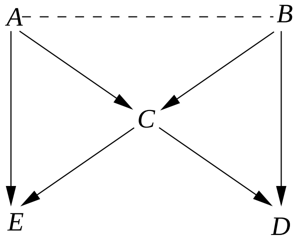

Figure 3:

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 Fig. 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 just separated) by a set S if, after conditioning on S, there is no open path between them. Thus in Fig. 1, $\{A, C\}$ separates E from B, but $\{C\}$ does not (because conditioning on C alone results in Fig. 2, in which E and B are connected via the open path A). In a DAG, $pa[X]$ 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 Fig. 1 $pa[E] = \{A, C\}$, which separates E from B but not from D.

Dependencies induced by conditioning on a set S can be read directly from the original graph using the criterion of d-separation, by tracing the original paths in the graph while testing whether colliders are, or have, descendants in S. The conditional dependencies are then illustrated in the original graph by drawing a circle around each C in S to denote the conditioning, then defining a path blocked by S if C is a noncollider on the path, or by a circle-free collider that does not have a circled desendant. Thus if we circle C in Fig. 1, it will completely 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 circled collider C, which is equivalent to having a dashed arc as in Fig. 2. Were we to circle F but not C, no open path would be completely blocked, but the collider C would again be opened by virtue of its circled descendant F, which is equivalent to having a dashed arc as in Fig. 3.

Selection Bias and Confounding

There is considerable variation in the literature in the usage of terms like "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 a nondirected open path between X and Y is a *biasing path* for the dependence of Y on X. The latter dependence is then *unbiased* for the effect of X on Y if the only open paths from X to Y are the directed paths. Next, consider a set of variables S that contains no effect (descendant) of X (including those descended through Y). 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.

The exclusion from S of descendants of X in these definitions arises first, because conditioning on X-descendants Z can partially block directed (causal) paths that are part of the effect of interest (if those descendants are intermediates or descendants of intermediates); and second, because conditioning on X descendants can unblock or create paths that are not part of the $X - Y$ effect, and thus create new bias. For example, biasing paths can be created when one conditions on a descendant Z of both X and Y. The resulting bias is called *Berksonian bias*, after its discoverer, Joseph Berkson.

Informally, confounding is a source of bias arising from causes of Y that are associated with but not affected by X. 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. Note that an unconfounded dependency may still be biased due to nondirected open paths that do not end in an arrow into Y (e.g., if Berksonian bias is present).

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 the directed paths. Consider again a set of variables S that contains no descendant of X. 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.

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 S contains no descendant of X and 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, hence the dependence of Y on X is unbiased whenever there is no open back-door path from X to Y;

2. If X is exogenous, the dependence of any Y on X is unbiased;

3. All confounders are ancestors of either X or of Y;

4. A back-door path is open if and only if it contains a common ancestor of X and Y;

5. 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 Fig. 2. Also, although $pa[X]$ always satisfies the back-door criterion and hence is sufficient in a DAG, it may be far from minimal sufficient. For example, in a DAG there is no confounding and hence no need for conditioning whenever X separates $pa[X]$ from Y (i.e., whenever the only open paths from $p[X]$ to Y are through X).

The terms "confounding" and "selection bias" have somewhat varying and overlapping usage. Epidemiologists typically refer to Berksonian bias as "selection bias," and some call any bias created by conditioning "selection bias." Nonetheless, some writers (especially in econometrics) use "selection bias" to refer to what epidemiologists call confounding. Indeed, Figs. 1 and 3 show how selection on a nonconfounder (F) can generate confounding. As a final caution, we note that the biases dealt with by the above concepts are only confounding and selection biases. Biases due to measurement error and model-form misspecification require further structure to describe.

Statistical Interpretations

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 $pa[X]$. A distribution compatible with a DAG thus can be reduced to a product of factors $P(x|pa[X])$, 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 the probability of treatment (propensity score) is fully determined by the parents of $X, pa[X]$. Roughly speaking, the factorization implies that a distribution compatible with a complete causal structure factorizes into the product of the propensity scores for each variable in the structure.

Suppose 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 conditioning 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 $X = x_1$ vs. $X = x0$ on Y when $S = s$ is defined as $P(y|x_1, s) - P(y|x_0, s)$, 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 $P(y|x_1, s)/P(y|x_0, s)$. A *standardized effect* is a difference or ratio of weighted averages of these stratum-specific $P(y|x, s)$ over S, using a common weighting distribution. 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").

The above form of standardized effect is identical to the forms derived under other causal models.When S is sufficient, some authors go so far as to identify the $P(y|x, s)$ with the distribution of potential outcomes given S. There have been objections to this identification on the grounds that not all variables in the graph can be manipulated, and that potential-outcome models do not apply to nonmanipulable variables. The objection loses force when X is an intervention variable, however. In that case, sufficiency of a set S implies that the potential-outcome distribution equals $\sum_s P(y|x, s)P(s)$, the risk of $Y = y$ given $X = x$ standardized to the S distribution.

Some Epidemiologic Applications

To check sufficiency and identify minimally sufficient sets of variables given a graph of the causal structure, one need only see 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 and eases teaching nonstatisticians algebraically difficult concepts.

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, 0 otherwise, as well as the study variables, 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.

Fig. 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 Fig. 4 this would be a claim that, upon stratifying on F, the $E - D$ association represents the direct effect of E on D. Fig. 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.

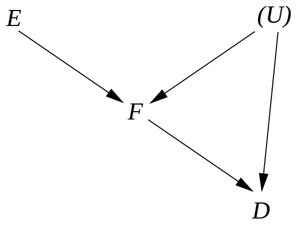

Figure 4:

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

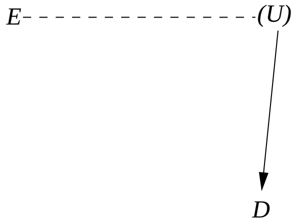

Figure 5:

The generation of biasing paths by conditioning on a collider or its descendant has been called "collider bias." Starting from a DAG, there are two distinct forms of this bias: Confounding induced in the conditional graph (Figs. 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 remove confounding; in contrast, no variable need appear that could be used to remove Berksonian bias.

Fig. 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. Subject to additional assumptions, one can put bounds on confounding of the $F - D$ association (and with more assumptions remove it entirely) through use of E as an *instrumental variable* (a variable associated with X and separated from Y by X).

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. A key assumption in certain "discovery" algorithms is a converse of compatibility called *faithfulness.*

A compatible distribution is *faithful* to or *perfectly compatible* with a given graph 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 contains no independencies other than those implied by graphical separation). A distribution is *stable* if there is a DAG to which it is faithful. Methods exist for constructing a distribution that is faithful to a given DAG. Methods also exist for constructing a minimal DAG compatible with a given distribution (minimal in that no arrow can be removed from the DAG without violating compatibility). 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 graphical models discussed here.

Whether or not one assumes faithfulness, the generality of graphical models is purchased with limitations on their informativeness. The nonparametric nature of the graphs implies that parametric concepts like effect modification cannot be displayed by the graphs (although the graphs still show whether the effects and hence their modification can be estimated from the given information). 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 Fig. 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. 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. As a consequence, there is often good reason to expect certain phenomena (such as the conditional $E - D$ confounding shown in Figs. 2, 3 and 5) to be small in epidemiologic examples. Thus, when quantitative information is used, graphical modeling becomes more a schematic adjunct than an alternative to causal modeling.

See also: Bias, Types of; Causation and Causal Inference; Confounding

Further Readings

Full technical details of causal diagrams and their relation to causal inference can be found in Pearl (2000) and Spirtes et al. (2001). Less technical reviews geared toward health scientists include Greenland et al. (1999), Greenland and Brumback (2002), Jewell (2004), and Glymour (2006).

General

Glymour, M.M. (2006). Using causal diagrams to understand common problems in social epidemiology. In: Oakes, J.M. and Kaufman J.S. (eds.). *Methods in Social Epidemiology.* San Francisco: Jossey-Bass, in press.

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

Greenland, S. and Brumback, B.A. (2002). An overview of relations among causal modelling methods. *International Journal of Epidemiology*, 31, 1030-1037.

Jewell, N.P. (2004). *Statistics for Epidemiology.* Boca Raton: Chapman and Hall/CRC, sec. 8.3.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, CA.

Pearl, J. (1995). Causal diagrams for empirical research (with discussion). *Biometrika*, 82, 669-710.

Pearl, J. (2000). *Causality.* New York: Cambridge University Press.

Pearl, J. (2001). Causal Inference in the Health Sciences: A Conceptual Introduction. *Health Services and Outcomes Research Methodology,*, 189-220.

Spirtes P., Glymour C., Scheines R. *Causation, Prediction, and Search*, 2nd ed. Cambridge, MA: MIT Press, 2001.

Methodologic Applications and Issues

Cole S., Hernán MA. (2002). Fallibility in estimating direct effects. *International Journal of Epidemiology* 31: 163-165.

Freedman D.A. and Humphreys P. (1999). Are there algorithms that discover causal structure? *Synthese*, 121, 29–54.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, 29, 722-729. (Erratum: 2000, 29, 1102)

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

Hernán, M.A., Hernandez-Diaz, S., Werler, M.M., Mitchell, A.A. (2002). Causal knowledge as a prerequisite for confounding evaluation. *American Journal of Epidemiology*, 155:176-184.

Hernán, M.A., Hernandez-Diaz, S., Robins, J.M. (2004). A structural approach to selection bias. *Epidemiology*, 15, 615-625.

Robins J.M. Data, design, and background knowledge in etiologic inference. *Epidemiology* 2001;12:313-320.

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