Supplement 2. Use of Directed Acyclic Graphs

Brian Sauer, Ph.D.
University of Utah School of Medicine, Salt Lake City, UT

Tyler J. VanderWeele, Ph.D.
Harvard School of Public Health, Boston, MA

Abstract

This supplement describes how counterfactual theory is used to define causal effects and the conditions in which observed data can be used to estimate counterfactual-based causal effects. Basic definitions and language used in causal graph theory are then presented. The graphical separation rules linking the causal assumptions encoded in a diagram to the statistical relations implied by the causal diagrams are then presented. The supplement concludes with a description of how Directed Acyclic Graphs (DAGs) can be used to select covariates for statistical adjustment, identify sources of bias, and support causal interpretation in comparative effectiveness studies.

Introduction

Under the rubric of structural equation modeling, causal diagrams were historically used to illustrate qualitative assumptions in linear equation systems. Judea Pearl extended the interpretation of causal diagrams to probability models, a development that has enabled the use of graph theory in probabilistic and counterfactual inference.[1] Epidemiologists then recognized that these diagrams could be used to illustrate sources of bias in epidemiological research,[2] and for this reason have recommended the use of causal graphs to illustrate sources of bias and to determine if the effect of interest can be identified from available data.[3-6]

This supplement begins with a brief overview of how counterfactual theory is used to define causal effects and of the conditions under which observed data can be used to estimate counterfactual-based causal effects. We then present the basic definitions and language used in causal graph theory. Next we describe the construction of causal diagrams and the graphical separation rules linking the causal assumptions encoded in a diagram to the statistical relations implied by the diagram. The supplement concludes with a description of how Directed Acyclic Graphs (DAGs) can be used to select covariates for statistical adjustment, identify sources of bias, and support causal interpretation in comparative effectiveness studies.

Estimating Causal Effects

The primary goal of nonexperimental comparative effectiveness research is to compare the effect of study treatments on the risk of specific outcomes in a target population. To determine if a treatment had a causal effect on an outcome of interest, we would like to compare individual-level outcomes under each treatment level. Unfortunately, an individual's outcome can only be observed under one treatment condition, which is often referred to as the factual outcome. Outcomes under treatment conditions not actually observed are referred to as counterfactual or potential outcomes.[7-8] Using counterfactual theory, we would say that a treatment had a causal effect on an individual's outcome if the outcome experienced would have been different under an alternative treatment level. For example, we would conclude that treatment A had a causal effect on the outcome Y if, say, an individual died 5 days after taking the drug (a=1), but would have remained alive on day 5 if he had not taken the medication (a=0). Due to the missing counterfactual data, causal effect measures cannot be directly computed for individual people without very strong assumptions. Nevertheless, average causal effects can be consistently estimated in randomized experiments and nonexperimental studies under certain assumptions.[7-8]

Assuming that we can simultaneously estimate the outcome risk for the entire population under different treatment conditions, then an average causal effect occurs when the outcome risk is not equal across levels of treatment. Using a dichotomous treatment (A) and outcome (Y) as the example, the causal effect in a population is the probability of the outcome occurring when the entire population is treated $Pr[Y^{a=1}=1]$ minus the probability of the outcome occurring when the entire population is untreated $Pr[Y^{a=0}=1]$. Populations, like individuals, cannot simultaneously receive different levels of treatment. We can, however, use observed data to draw inferences about the probability distributions or expectations over a population of counterfactual variables. One of the important assumptions required for using only observed data (factual data) to estimate average causal effects is exchangeability.

In an *ideal* randomized experiment, treatment assignment is independent of the counterfactual outcomes, and therefore the two groups are exchangeable.[7, 9] This means that the risk of experiencing the outcome in the two groups at the time of treatment assignment is equal to the risk in the full population. The equivalency to the full population allows us to use the observed data from the treated group to estimate what the treatment effect would have been if the entire population was treated, and it also allows us to use the observed data from the untreated group to estimate the effect of no treatment in the full population. In addition, because the outcome risks in the subpopulations are equivalent at the time treatment is assigned, the observed risk difference between the treatment groups can be attributed to treatment effects.[10] In an ideal randomized trial, the outcome experience had the entire population been treated ($Pr[Y^{a=1}=1]$) is equal to the probability of the outcome occurring in the subset of the population who received treatment ($Pr[Y=1|A=1]$), and the same holds for the untreated group. Using the risk difference scale, this means that the conditional risk difference can be used to estimate the marginal causal risk difference ($Pr[Y=1|A=1]$ - $Pr[Y=1|A=0]$) = ($Pr[Y^{a=1}=1]$ - ($Pr[Y^{a=0}=1]$).

In nonexperimental studies, marginal exchangeability can rarely be assumed, since patients and providers typically select treatments based on their belief about the risk of specific outcomes. In this case marginal exchangeability does not hold, but exchangeability may hold within levels of risk factors pertaining to the outcome. Causal inference from nonexperimental data is based on the critical assumption that within levels of important risk factors, treatment assignment is effectively randomized. This assumption is also referred to as "conditional exchangeability," "conditional unconfoundedness," or the assumption of "conditionally ignorable treatment assignment."[8, 10] When we assume that treatment was randomly assigned conditional on a set of covariates, causal inference for nonexperimental comparative effectiveness studies requires some form of covariate adjustment.

The question then concerns the adjustments that must be made in order to generate conditional exchangeability and avoid bias. DAGs have been found to be particularly helpful in diagnosing sources of bias and helping investigators select a set of covariates that allow the estimation of causal effects from observed data. Using DAG theory, confounding bias can be characterized as an unblocked "backdoor" path from the treatment to the outcome. The next section presents terminology for DAGs and their utility in selecting covariates for statistical adjustment.

DAG Terminology

DAGs are used to encode researchers' a priori assumptions about the relationships between and among variables in causal structures. DAGs contain directed *edges* (arrows), linking *nodes* (variables), and their *paths*. A path is an unbroken sequence of distinct nodes connected by edges; a directed path is a path that follows the edges in the direction indicated by the arrows, such as the path from A to C (A→B→C). An undirected path does not follow the direction of the arrows, such as the following A to C path (A→B→C). Kinship terms are often used to represent relationships within a path. If there exists a directed path from A to C, then A is an *ancestor* of C and C is a *descendent* of A. Using the directed path example of A→B→C,

A is a *direct cause* or *parent* of B, and B is a *child* of A and parent of C, while A is considered an *indirect cause* or *ancestor* of C. The node B lies on the causal pathway between A and C and is considered an *intermediate* or *mediator* variable on the directed path. DAGs are acyclic since no node can have an arrow pointing to itself, and all edges must be directed (contain arrows).[2] In other words, no directed path from any node to itself is allowed. These rules enforce the understanding that causes must precede their effects. Mutual causation is handled in DAGs by including a time component, which allows A to cause B at time (t) and B to cause A at some later time ($t+1$).

The first step in creating a causal DAG is to diagram the investigators' understanding of the relationships and dependencies among variables. Construction of DAGs should not be limited to measured variables from available data; they must be constructed independent of available data and of background knowledge of the causal network linking treatment to the outcome. The most important aspect of constructing a causal DAG is to include on the DAG any common cause of any other two variables on the DAG. Variables that only causally influence one other variable (exogenous variables) may be included or omitted from the DAG, but common causes must be included for the DAG to be considered causal. The absence of any path between two nodes in a DAG indicates that the variables are not causally related (i.e., that manipulation of one variable does not cause a change in the value of the other variable). Investigators may not agree on a single DAG to represent a complex clinical question; when this occurs, multiple DAGs may be constructed and statistical associations observed from available data may be used to evaluate the consistency of observed probability distributions with the proposed DAGs. Statistical analyses may be undertaken as informed by different DAGs, and the results can be compared.

Figure S2.1 is a modified DAG illustrating a highly simplified hypothetical study, described in chapter 7, to compare rates of acute liver failure between new users of calcium channel blockers (CCBs) and diuretics.
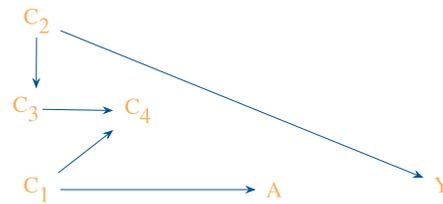

Figure S2.1 Hypothetical DAG illustrating causal relationships among formulary policy (C_1) and treatment with a CCB (A) and treatment for eretile dysfunction (C_4). Alcohiol abuse (C_2) influences impotence (C_3), which influences treatment of erectile dsyfuncution (C_4) and is a cause of acute liver disease (Y). In this example there is no effect of antihypertensive treatment, that is, treatment with a CCB (A), on liver disease.

Causal diagrams like Figure S2.1 can be used to express causal assumptions and the statistical implications of these assumptions.[11-12]

Independence Relationships

DAGs can be used to infer dependence and conditional independence relationships if the causal structure represented by the graph is correct. The rules linking the structure of the graph to statistical independence are called the d-separation criteria and are stated in terms of blocked and unblocked paths.[2] To discuss blocked and unblocked paths, we need one more graphical concept, that of a collider. A node is said to be a collider on a specific path if it is a common effect of two variables on that path (i.e., when both the preceding and subsequent nodes have directed edges going into the collider node). In Figure S2.1, C_4 is a collider on the path $A \leftarrow C_1 \rightarrow C_4 \leftarrow C_3 \leftarrow C_2 \rightarrow Y$. Note, however, that whether a variable is a collider or not is relative to the path. C_4 is not a collider on the path $C_4 \leftarrow C_3 \leftarrow C_2 \rightarrow Y$.

We can now define blocked paths. A path from a node A to a node Y is unconditionally blocked if there is a collider on the path from A to Y (e.g., Figure S2.2). A path from a node A to a node Y is said to be blocked conditional (e.g., when adjusting) on a set of variables Z if either there is a variable in Z on the path that is not a collider or if there is a collider on a path such that neither the collider nor any of its descendants are in Z. Otherwise, the path is said to be unblocked or open. Two paths between A and Y exist in Figure S2.2. The path $A \leftarrow C_1 \rightarrow C_4 \rightarrow C_5 \rightarrow Y$ is an open path,

while the $A \leftarrow C_1 \rightarrow C_4 \leftarrow C_3 \leftarrow C_2 \rightarrow Y$ path is closed due to the collider C_4. Adjustment for C_4 or C_5 will close the $A \leftarrow C_1 \rightarrow C_4 \rightarrow C_5 \rightarrow Y$ path but open a backdoor path on the $A \leftarrow C_1 \rightarrow C_4 \leftarrow C_3 \leftarrow C_2 \rightarrow Y$ pathway by inducing an association between C_1 and C_3. Adjustment for C_1 alone will close the open $A \leftarrow C_1 \rightarrow C_4 \rightarrow C_5 \rightarrow Y$ path and not alter the $A \leftarrow C_1 \rightarrow C_4 \leftarrow C_3 \leftarrow C_2 \rightarrow Y$ path, which is closed due to the collider.
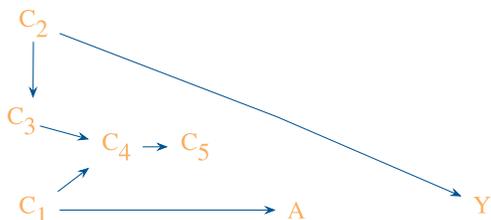

Figure S2.2 Hypothetical DAG used to illustrate the open backdoor path rule. Adjustment for C_4 or C_5 will open the $A \leftarrow C_1 \rightarrow C_4 \leftarrow C_3 \leftarrow C_2 \rightarrow Y$ path. Adjustment for C1 will close the open $A \leftarrow C_1 \rightarrow C_4 \rightarrow C_5 \rightarrow Y$ path.

Blocked paths correspond to independence; unblocked paths to association. More formally, we say that a node A and a node Y are d-separated conditional on Z if all paths from A to Y are blocked conditional on Z. If a DAG correctly describes the causal structures, then it follows that if A and Y are d-separated conditional on Z, then A and Y are conditionally independent given Z. This is sometimes referred to as the d-separation criterion. On the other hand, variables that are marginally independent but have a common effect become conditionally dependent when statistically adjusting the common effect. Adjusting for such colliders is said to open up backdoor paths and induce conditional associations. A stylized example used to illustrate this concept describes two ways in which the pavement (X_3) can be wet—the sprinkler system (X_1) is on or it is raining outside (X_2).[11] One assumes that the owners of the sprinkler system watered their lawn based on a preprogrammed schedule, making use of sprinklers unassociated with rain. Suppose you had a data table with data on X_1, X_2 and X_3 during the past year. If you were to evaluate the association between X_1 and X_2, you would find that X_1 does not predict X_2 and X_2 does not predict X_1. Now suppose you only use data where the concrete is

wet and reevaluate the association between X_1 and X_2. By conditioning on the concrete being wet $(X_3 = 1)$, dependence is established between the sprinklers being on and rain that did not previously exist. For example, if we know the concrete is wet and we also know the sprinklers are not on, then we can predict that it must be raining. Conditioning on a collider by either statistical adjustment or selection into the study can generate unintended consequences and bias the effect estimate.

Using DAGs To Select Covariates and Diagnose Bias

In a nonexperimental setting, the goal of covariate selection is to remove confounding by covariate selection. As described in chapter 7, intermediate, collider, and instrumental variables may behave statistically like confounders. For this reason, background knowledge is required to distinguish confounders for statistical adjustment. The most important result relating conditional exchangeability to causal diagrams is Pearl's backdoor path adjustment theorem, which provides a simple graphical test that investigators can use to determine whether the effect of A on Y is confounded. A set of variables, Z, satisfies the backdoor criterion relative to the treatment A and outcome Y in a DAG if no node in Z is a descendant of A and Z blocks every path between A and Y that begins with an arrow into A. The backdoor path adjustment theorem states that if Z satisfies the backdoor path criterion with respect to A and Y then the treatment groups are exchangeable conditional on Z.[1]

Using the backdoor path adjustment theorem, we can see the close connection between backdoor paths and common causes. Figure S2.3 indicates that treatment (A) and outcome (Y) have a common cause (C_4). The backdoor path from A to Y is open and confounding is present unless C_4 is statistically adjusted. We will represent conditioning on a variable by placing a square around the node, as illustrated in Figure S2.3. Unfortunately, adjustment for C_4 opens a backdoor path from A to Y through C_1, C_4, C_3, and C_2, resulting in bias, unless additional adjustment is

made for C_1, C_2, or C_3, or any combination of these. The key to ensuring conditional exchangeability is to measure and condition on variables needed to block all backdoor paths between the treatment and outcome (i.e., to condition on a sufficient set of confounders). When the effect of A on Y is unconfounded given a set of variables Z, we can then estimate the average causal effect described above using observed conditional probabilities $(\Pr[Y=1|A=1, Z=z] - \Pr[Y=1|A=0, Z=z]) = (\Pr[Y^{a=1}=1|Z=z] - \Pr[Y^{a=0}=1|Z=z])$.
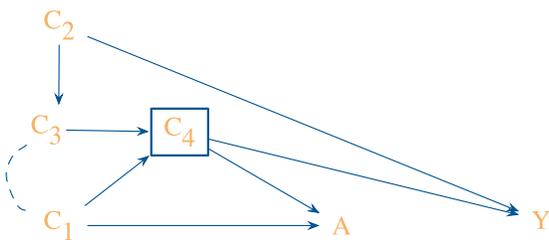

Figure S2.3 DAG illustrating causal relationships among formulary policy (C_1) and treatment with a CCB (A) and treatment for erectile dysfunction (C_4). Alcohol abuse (C_2) influences impotence (C_3), which influences treatment of erectile dysfunction (C_4) and is a cause of acute liver disease (Y). In this example C_4 is a confounder and collider. Adjustment of C_4 is additional to adjustment for at least one other variable on the open C_{1-3} pathways.

Using DAGs To Diagnose Selection Bias

The previous section described the use of DAGs to remove confounding, thereby enabling the estimation of average causal effects using observed patient responses to treatment. This section describes the use of DAGs to diagnose bias that results from selection into a study. Selection bias results when the estimated causal effect is different in the subset of the population being evaluated, when the goal is to make an inference to the full population. Selection bias occurs when the risk for the outcome in the population being studied is different from the risk in the target population, a situation that can happen when study participants are not representative of the target population. Various causes of selection bias have been described as healthy-worker bias, volunteer bias, selection of controls into case-control studies, differential loss-to-followup, and nonresponse.

In the previous section, we described a type of selection bias that occurs when conditioning on a collider variable. We called this situation collider stratification bias. This bias occurs from estimating the average causal effect within "selected" stratum, then averaging across strata. It turns out that the basic structure of selection bias is the same as collider stratification bias, which has been described as conditioning on a common effect of two other variables.[6] In the following section, we provide an example of how conditioning on a common effect can result from differential loss to followup. Please review the paper by Hernán and colleagues titled "A Structural Approach to Selection Bias" for a more complete discussion of other forms of selection bias.[6]

Selection bias is a result of conditioning on a common effect of two variables. To simplify, consider a randomized trial of antihypertensive treatments (CCB or other) and the outcome of acute liver disease (Y). The DAG in Figure S2.4 indicates that A is not causally associated with Y, but we would expect an association between A and Y conditional on S (selection) even though A does not cause Y. Assume that patients initiated on CCB have a higher rate of experiencing adverse drug effects and are more likely to drop out of the study (S=1) as represented from the arrow from A to S. Further assume that patients who abuse alcohol (C=1) are more likely to drop out as well. The square around S indicates that the analysis is restricted to individuals who did not drop out of the study.

Due to the random assignment of A, the variables A and C are marginally independent, but become conditionally dependent when selecting only subjects who remained in the study (i.e., those who did not drop out). Knowing that a study subject was an alcohol abuser but remained in the study suggests that she did not experience adverse effects of therapy. Restricting this analysis to subjects who did not drop out will result in patients treated with CCB having a lower proportion of alcohol abuse, thus making CCBs appear to be protective against acute liver failure when no causal association exists. This conditional dependence opens a pathway from A to Y through C thus biasing the observed risk difference from the counterfactual risk difference and resulting in selection bias.
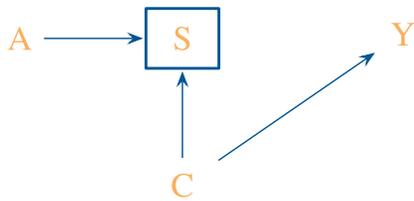
181

Figure S2.4. DAG illustrating selection bias. Treatment (A) is randomized. Subjects randomized to CCBs (A=1) are more likely to drop out due to adverse drug effects. Subjects with alcohol abuse (C=1) are more likely to drop out of the study and they are also more likely to experience acute liver failure (Y=1). Conditioning on selection (retention in study) (S=1) induces as association between A and C, which results in an open biasing pathway between A and Y.

There are situations where the causal risk estimate can be recovered from a design affected by selection bias. A technique called inverse probability weighting that generates a pseudopopulation where all subjects remained in the study can, under certain assumptions, be used to estimate the average causal effect in the entire target population. Inverse probability weighting is based on assigning a weight to each selected subject so that she accounts in the analysis not only for herself but also for those with similar characteristics (i.e., those with the same values of C and A) in subjects who were not selected.[6] The effect measure based on the pseudopopulation, in contrast to that based on the selected population, is unaffected by selection bias provided that the outcome of the uncensored subjects truly represents the unobserved outcomes of the censored subjects. This provision will be satisfied if the probability of selection is calculated conditional on A and all other factors that independently predict both selection and the outcome. However, this is an untestable assumption and one must carefully consider influences of discontinuation and the outcome when attempting to statistically address selection bias.

Conclusion

This supplement described the use of DAGs to identify sources of bias in nonexperimental comparative effectiveness research. The goal of covariate selection is to generate conditional exchangeability, thereby allowing unbiased

causal effect estimates within strata of covariates that are then pooled in some manner to generate unbiased average causal effects. The challenge of nonexperimental research is choosing a set of covariates that removes confounding bias and does not inadvertently generate other sources of bias. A confounder is typically considered a common cause of treatment and outcome, and DAG theory conceptualizes confounding as an open pathway between treatment and outcome. Confounders, unfortunately, cannot be selected based on statistical associations alone because some types of bias-inducing variables behave statistically like confounders. A common effect of two variables on a backdoor pathway is considered a collider. Colliders behave statistically like confounders, but pathways that include colliders are considered closed and do not bias the targeted effect estimate. Adjustment for colliders opens up additional pathways that can generate bias if necessary variables on the newly opened pathway are not appropriately adjusted.

Conditioning on the common effect of two variables (i.e., colliders) turns out to be the structural explanation for all types of selection bias. Selection bias occurs when participation in the study though volunteerism, design, adherence to treatment, or followup is influenced by the treatment and either the outcome or risk factors for the outcome. Some forms of selection bias, such as differential loss to followup, can be corrected by statistical techniques that analyze a pseudopopulation based on the subpopulation that was not lost to followup.

The use of DAGs can help researchers clarify and discuss their beliefs about the underlying data generating process, which can in turn aid the interpretation of the statistical associations observed in the data. Developing DAGs is not always easy and may require a heuristic approach, where assumptions are tested by observed statistical association and revised. A disciplined approach to developing DAGs may be useful for communicating findings and providing rationale for covariate selection. As discussed in chapter 7, there are often situations where a complete understanding of the causal network linking treatment to outcome is unknown. Empirical variable selection techniques may be employed to identify potential confounders for consideration.

In addition, we described methods for selecting covariates based on incomplete knowledge of the causal structure. In this case, simplifying rules, such as selecting all direct causes of treatment and/ or outcome may, in certain circumstances, be a good technique for removing confounding when the full causal structure is unknown.[13] Familiarity with DAG theory will improve the investigators' understanding of the logic and principles behind covariate selection for nonexperimental CER. Furthermore, use of DAGs standardizes the language for covariate selection, thus improving communication and clarity within the field and among investigators.

Checklist: Guidance and key considerations for DAG development and use in CER protocols		
Guidance	**Key Considerations**	**Check**
Develop a simplified DAG to illustrate concerns about bias.	– Use a DAG to illustrate and communicate known sources of bias, such as important well known confounders and causes of selection bias.	☐
Develop complete DAG(s) to identify a minimal set of covariates.	– Construction of DAGs should not be limited to measured variables from available data; they must be constructed independent of available data. – The most important aspect of constructing a causal DAG is to include on the DAG any common cause of any other two variables on the DAG. – Variables that only causally influence one other variable (exogenous variables) may be included or omitted from the DAG, but common causes must be included for the DAG to be considered causal. – Identify a minimal set of covariates that blocks all backdoor paths and does not inadvertently open closed pathways by conditioning on colliders or descendants.	☐

183

References

1. Pearl J. Causal inference from indirect experiments. Artificial intelligence in medicine. 1995;7:561-82.

2. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10:37-48.

3. VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. Am J Epidemiol. 2007;166:1096-104.

4. Robins JM. Data, design, and background knowledge in etiologic inference. Epidemiology. 2001;12:313-20.

5. Hernan MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002;155:176-84.

6. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15:615-25.

7. Hernan MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health. 2004;58:265-71.

8. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. J Am Stat Assoc. 2005;100:10.

9.   Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol. 1986;15:413-9.

10. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. 2006;60:578-86.

11. Pearl J. Section 1.5: Causal Versus Statistical Terminology. In: Causality: Models, Reasoning and Inference. New York: Cambridge University Press; 2000.

12. Glymour MM, Weuve J, Berkman LF, et al. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. Am J Epidemiol. 2005;162:267-78.

13. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. Biometrics. 2011;67: 1406-13.

184