

Introduction to Directed Acyclic Graphs (DAGs) for Causal Inference

Felix Elwert
University of Wisconsin-Madison
elwert@wisc.edu

I gratefully acknowledge the support of the SAMSI Program, *Data Science in the Social and Behavioral Sciences*, sponsored by the National Science Foundation, DMS-1929298 and DMS-1638521. I thank Elan Segarra for assistance.

Suggested Citations for This Workshop

Most material in this slide deck is drawn from these three papers:

- Elwert. 2013 “Graphical Causal Models.” Pp. 245-273 in S. Morgan (ed.), *Handbook of Causal Analysis for Social Research*. Springer.
- Elwert and Winship. 2014. “Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable.” *Annual Review of Sociology* 40: 31-54.
- Keele, Stevenson, Elwert. 2019. “The Causal Interpretation of Estimated Associations in Regression Models.” *Political Science Research and Methods* 8:1-13.

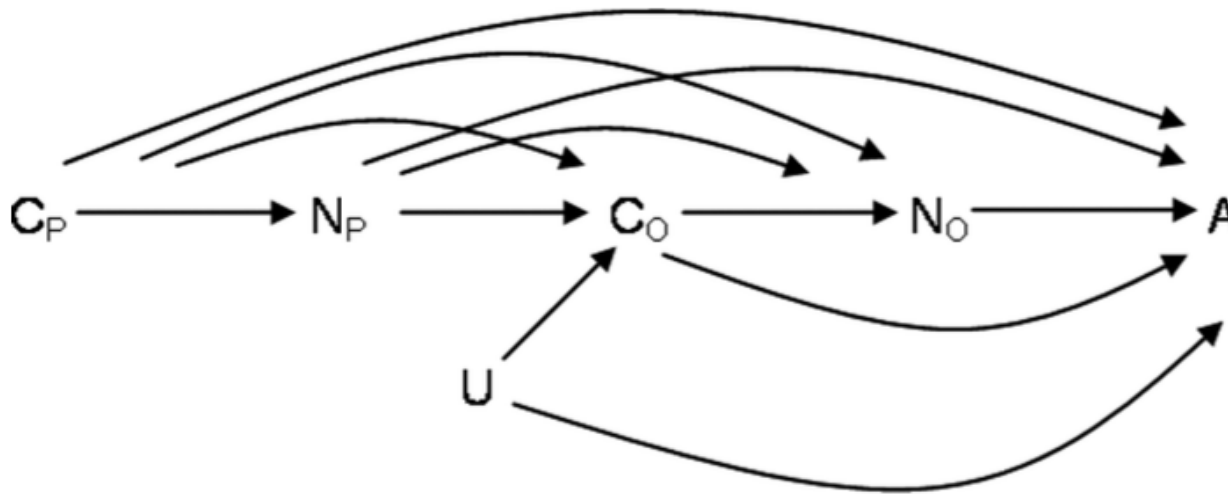
Comprehensive Technical References

You'll find comprehensive formal treatments here:

- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4), 669-710.
- Pearl, J. (2009). *Causality*. Cambridge University Press, second edition.
- Maathuis, M. H., Drton, M., Lauritzen, S., and Wainwright, M., editors (2018). *Handbook of Graphical Models*. CRC Press, first edition.

For additional works cited in this workshop, see the **Reference** section at the end of this slide deck.

What's a Directed Acyclic Graph?



This is a DAG!

- Developed by Judea Pearl (1988, 1995, 2009) and others in CS.
- Origins in structural equation modeling (social sciences, [Wright 1921]), and Bayesian networks (computer science, Pearl 1985).
- Compatible with Rubin's (1974) potential-outcomes framework of causality (Richardson & Robins 2013)

Overview: Four Main Uses

This workshop is structured to introduce the four main uses of DAGs.

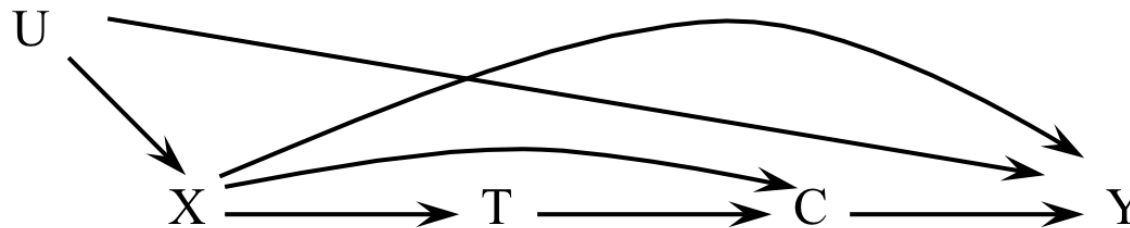
1. Graphically notate the assumed data-generating process (DGP)
2. Infer association from causation (d-separation)
3. Infer causation from association (adjustment criterion)
4. Understand thorny problems
 - i. Collider bias (selection)
 - ii. Interpreting regression coefficients
 - iii. Mediation analysis
 - iv. Instrumental variables

1. DAGs Notate the Assumed DGP

Elements—What’s in a DAG?

DAGs are pictures that encode the analyst’s qualitative causal (“structural”) assumptions about the data generating process (DGP), specifically, which variables cause what other variables.

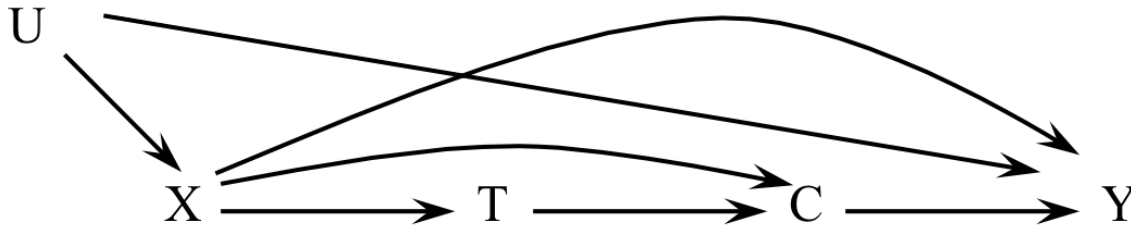
“DAGs are pictures of how the world works.”



1. Nodes represent variables (observed and unobserved)
2. Arrows (directed edges) possible direct causal effects
3. Missing arrows represent absent direct causal effects

Missing arrows are called “exclusion restrictions.” Without exclusions, one cannot point-identify causal effects.

Illustration: A Model of Smoking and Mortality



This DAG captures the following qualitative DGP:

Smoking, T , causes cancer, C : $T \rightarrow C$

Cancer, C , causes mortality, Y : $C \rightarrow Y$

Poverty, X , causes smoking, cancer, and mortality: $X \rightarrow T, X \rightarrow C, X \rightarrow Y$

Childhood accidents, U , cause poverty and mortality: $U \rightarrow X, U \rightarrow Y$

The DAG also includes several exclusion restrictions, e.g.,

Smoking causes mortality only via cancer: No arrow $T \rightarrow Y$

Remember: All causal claims made from data are relative to the assumed DGP.

If you do not believe this DGP, you should change it.

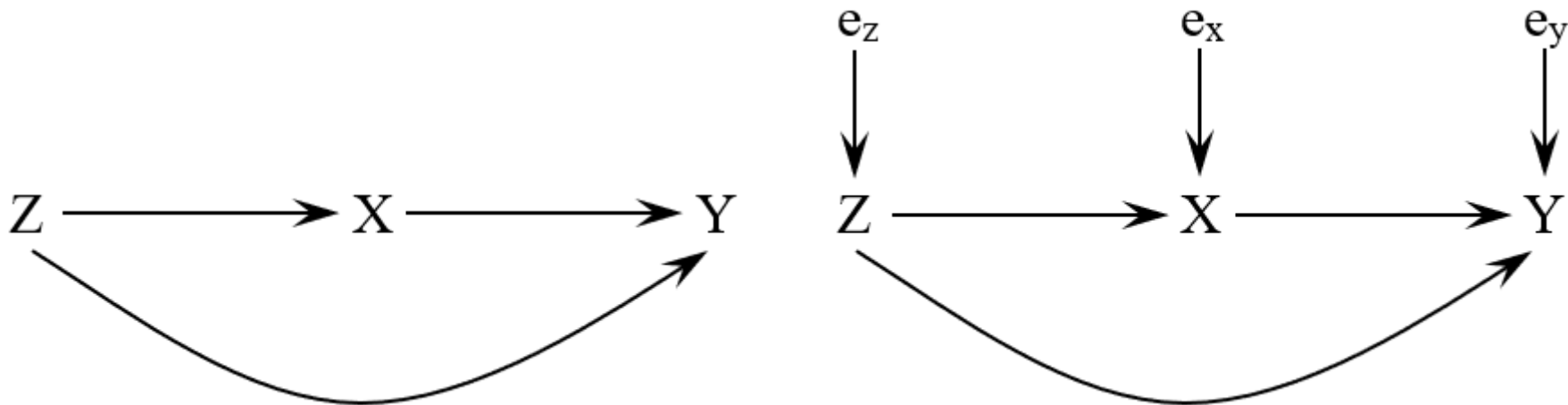
Idiosyncratic Structural Error Terms

Most variables have lots of causes. The unmeasured causes affecting each variable are called the “structural error term” of the variable.

An “idiosyncratic” structural error term includes all unobserved variables that only affect one measured variable in your graph.

By convention, the idiosyncratic error terms of the observed variables are not drawn in the DAG.

Here are two equivalent graphs, once drawn without, and once with, the idiosyncratic error terms. These two graphs represent the same DGP.

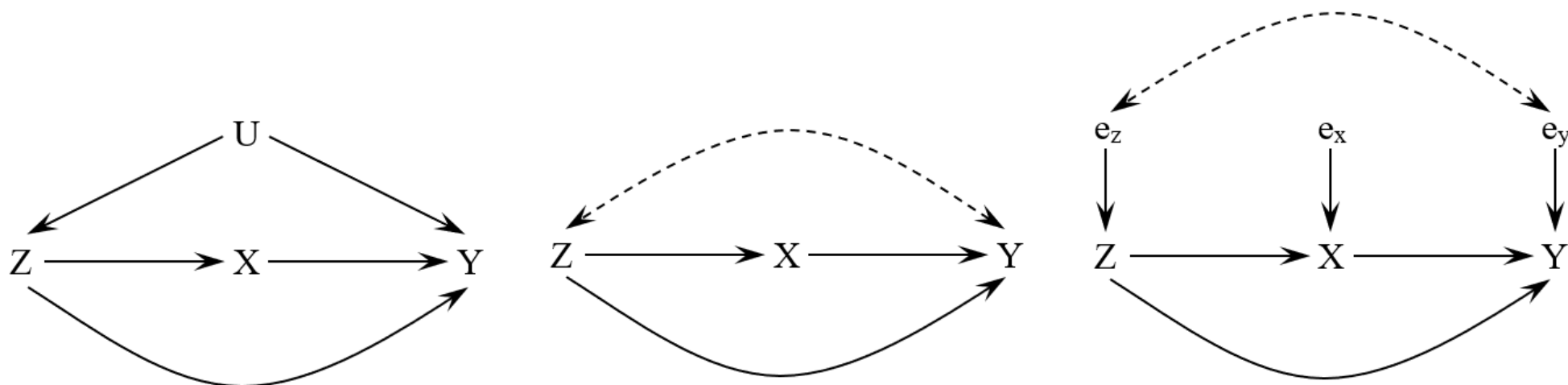


Correlated Error Terms

Sometimes, some *unmeasured* variable, U , may affect two or more variables in your DGP, e.g., Z and Y . Since U is thus part of the error term of both variables, we say that Z and Y have “correlated errors.”

Correlated errors always need to be drawn explicitly!

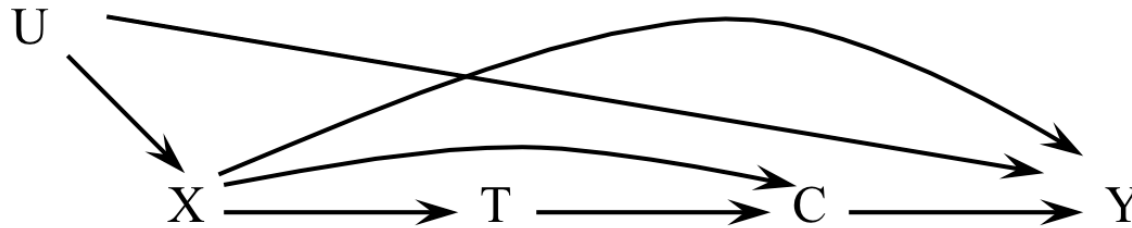
Here are three equivalent ways of drawing correlated error terms between X and Y . These three graphs represent the same DGP.



Causal DAGs

In what follows, we generally must assume that the DAG is a so-called “causal DAG.”

Definition: A causal DAG includes all observed and unobserved common (i.e., shared) causes of any two (or more) variables already included in the DAG.



E.g., if this DAG is to be “causal,” then we are asserting, among other things, that cancer, C , and mortality, Y , share no observed or unobserved common causes besides poverty, X .

Clearly, the assumption that your DAG is causal is a very strong, and frequently problematic, assumption. Causal inference requires strong assumptions.

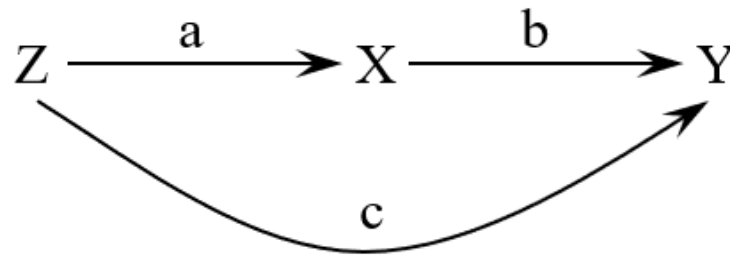
To proceed, we henceforth assume that all DAGs are causal.

DAGs vs. Path Models

Path models are a (very) special case of DAGs. The difference is that path models assume that the data are generated by a linear model, in which the causal effect of one variable on another is the exact same for every individual (“homogeneity”).

In a path model, the strength of the causal effect of each arrow is summarized by a single number, the so-called path coefficient.

Path coefficients¹ represent causal effects in the regression metric.



For example, this graph says that changing Z by one unit causes an increase in X by a units, and that this effect is the same for all individuals.

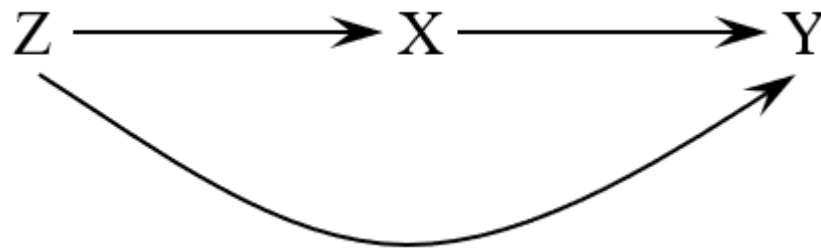
Furthermore, path analysis typically assumes that the nodes are jointly normal.

¹ “Path coefficients” is a misnomer—they are *parameters*, i.e., the structural/causal effects in the DGP.

DAGs are Non-parametric

In contrast to path models, DAGs are completely non-parametric objects.

1. No distributional assumptions (nodes can have any distribution)
2. No functional form assumptions (arrows can be linear or non-linear, effects may be homogenous or heterogeneous)

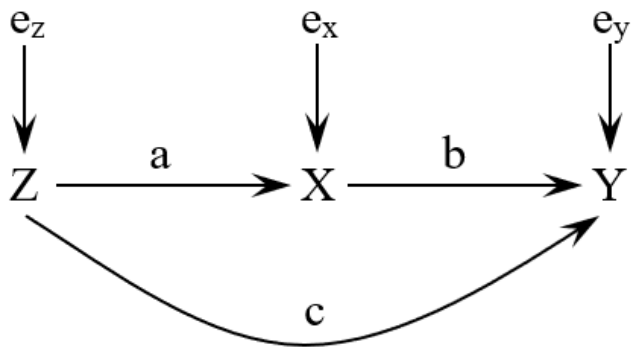


This makes DAGs a very general notation for the *qualitative* causal structure of the assumed DGP.

This generality guards against fallacious conclusions when distributional or functional form assumptions are ill justified (as they often are).

Graphs Represent Structural Equations

With all that, we have defined path models/DAGs as visual representations of the structural equations that are assumed to have generated the data.



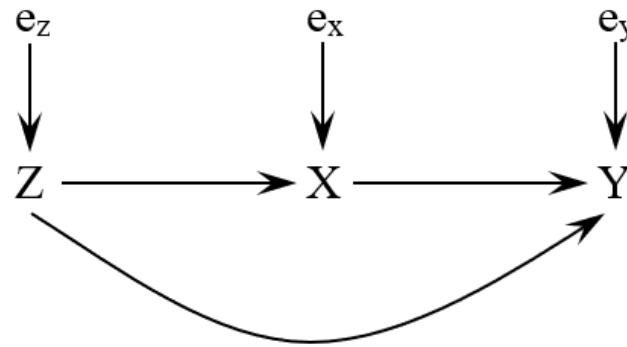
Linear path model

$$Z_i = e_{iz}$$

$$X_i = p + aZ_i + e_{ix}$$

$$Y_i = g + cZ_i + bX_i + e_{iy}$$

NB: $f_V(pa_V)$ says that V is generated (caused) by some (likely unknown) function, $f_V()$, that takes the parents of V (pa_V , all variables with a direct arrow into V) as inputs; $f_V()$ places no constraints on functional form and may include interactions.



Non-parametric structural equation model (NPSEM)

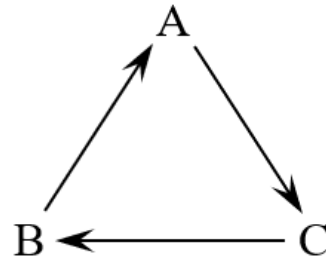
$$Z_i = e_{iz}$$

$$X_i = f_{iX}(Z_i, e_{ix})$$

$$Y_i = f_{iY}(Z_i, X_i, e_{iy})$$

DAGs are Acyclic = Contain No Directed Cycles

DAGs are “acyclic” in the sense that they don’t contain directed cycles of arrows.



Reason: Causes must occur before their effects. A directed cycle would say that the future causes the past => nonsense.

Apparent counterexamples are usually resolved by articulating the DAG in time (e.g., by adding time subscripts to multiple states of a variable).

There is also theory for cyclic, partially direct, and even undirected graphs. Naturally, such graphs support weaker conclusions, because they make weaker assumptions. See Pearl (2009) and Maathuis et al. (2018).

2. Inferring Association from Causation (d-separation)

Task: Inferring Association from Causation

Having notated the causal structure of the DGP graphically, we can now infer *what associations should exist* (be observable) in the data *if the graph represented the true DGP*.

This is known as “deriving the associational implications of a model.”

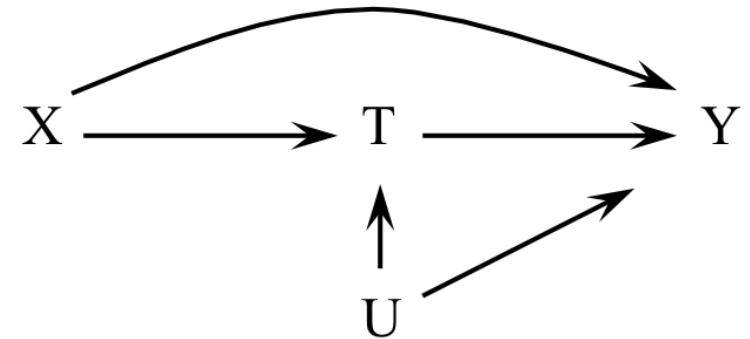
In other words, we ask: if this graph captures the process that generated the data, what variables should be associated with each other, and which ones should be independent?

In this section, we’ll build up to understanding the engine that connects the causal assumptions of the DAG to the statistical associations and independences in observed data and hence drives DAG-world: d-separation (Pearl 1988).

Outlook

Simple rules connect the causal assumptions in the DAG to statistical associations in the data. The logic goes like this:

1. The causal effects in a DGP give rise to observable associations in data.
2. All associations flow along open paths. By contrast, closed paths do not transmit association.
3. The three rules of association fully determine whether a path is open or closed.



In the following slides, we define these terms (path, open, closed) in order to lead up to the most important technical concept in DAG world: d-separation, which consolidates the three rules of association. See Pearl (2009) for formal details.

Paths: Causal and Non-causal

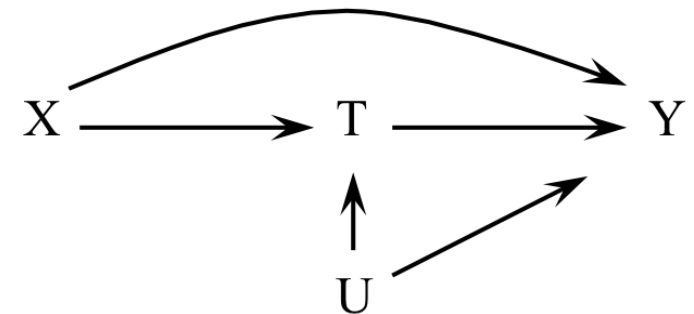
Definition: A path between two variables is a non-self-intersecting sequence of adjacent edges.

Meaning: the direction of the arrows does not matter; a path may not contain cycles (i.e., touch any variable more than once).

Every path between a treatment, T , and an outcome, Y , is either causal or a non-causal.

Definition: A causal path is a path in which all arrows point away from T and toward Y .

- The union of causal paths comprises the total causal effect.

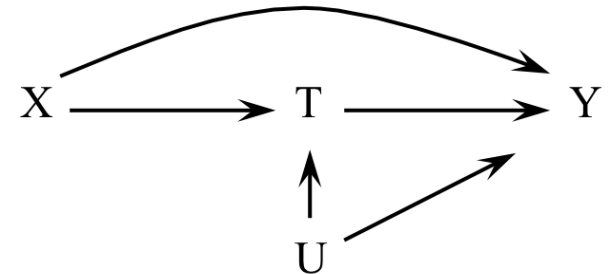


Definition: A non-causal path is path from T to Y in which at least one arrow points against the flow of time.

- Non-causal paths may carry spurious association.

Exercises

List the causal and non-causal paths from T to Y .



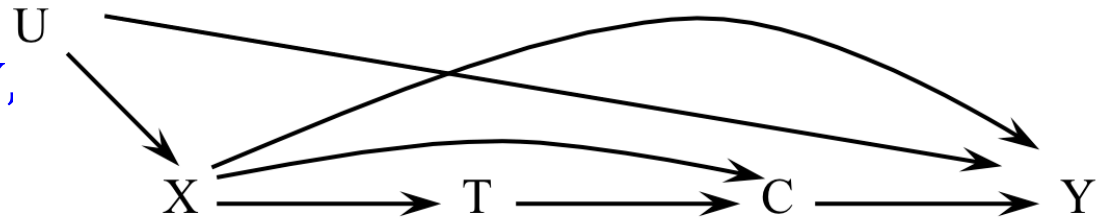
Causal: $T \rightarrow Y$

Non-causal: $T \leftarrow U \rightarrow Y$ and $T \leftarrow X \rightarrow Y$

List the causal and non-causal paths from X to Y .

Causal: $X \rightarrow Y$, $X \rightarrow C \rightarrow Y$,
 $X \rightarrow T \rightarrow C \rightarrow Y$

Non-causal: $X \leftarrow U \rightarrow Y$



Definition: A non-causal path between treatment, T , and outcome, Y , that begins with an arrow into T , $\rightarrow T$, is called a backdoor path.

Collider Variables Are Key

Collider variables are key for working with DAGs.

Definition: When two arrows on a given path point directly into a variable on the path, then that variable is a collider variable on the path.

Fact: On any given path, any one variable is either a collider or a non-collider.

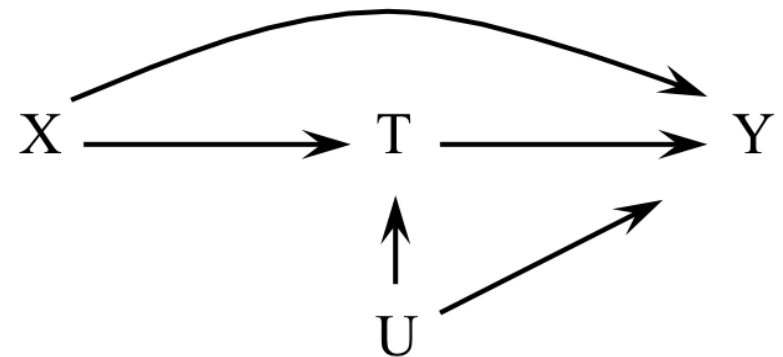
Ex: T is a collider on the path

$$X \rightarrow T \leftarrow U \rightarrow Y.$$

T is not a collider on the path

$$X \rightarrow T \rightarrow Y.$$

⇒ Colliders are path-specific!



Definition: A variable that is not a collider on a path is a non-collider on the path.

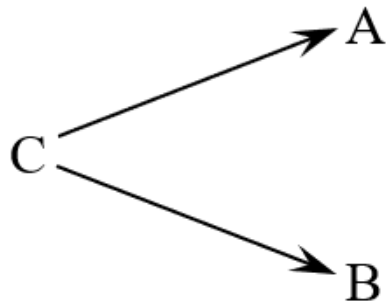
Three Rules of Association

Now let's ask which paths transmit, and which don't transmit, association. We start with the primitives. It turns out that all marginal and conditional associations originate from 3 causal structures:



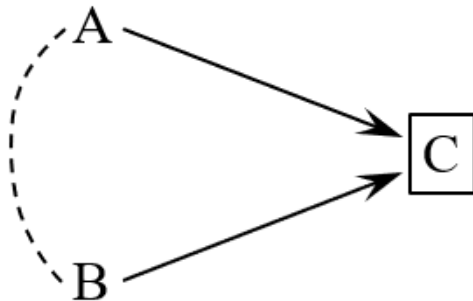
(1) Direct and indirect causation

$$A \nsim B \text{ and } A \perp B|C$$



(2) Common cause confounding

$$A \nsim B \text{ and } A \perp B|C$$



(3) Conditioning on a common effect ("collider"): Selection

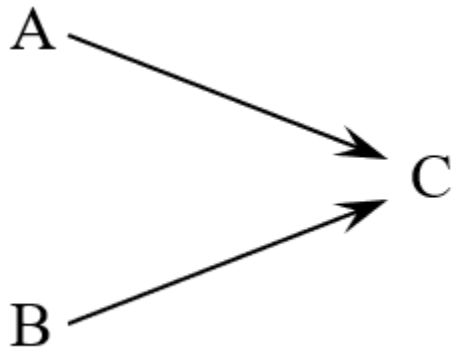
$$A \perp B \text{ and } A \nsim B|C$$

----- : non-causal (spurious) association

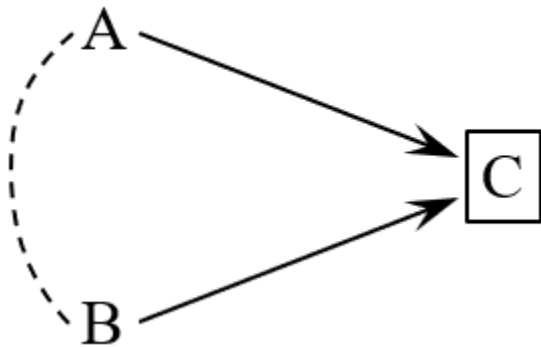
□ : conditioning

Conditioning on a Collider

Notice: No causal effect of A on B (or vice versa) and no confounding

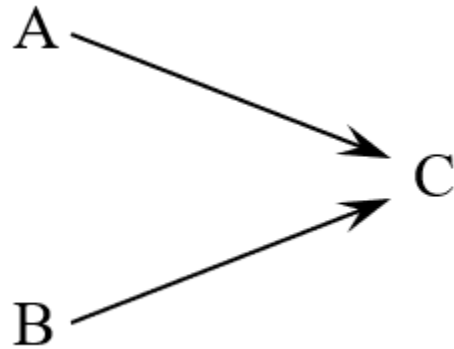


$A \perp B$: Marginally independent



$A \not\perp B|C$: Conditionally dependent

Conditioning on a Collider: Examples

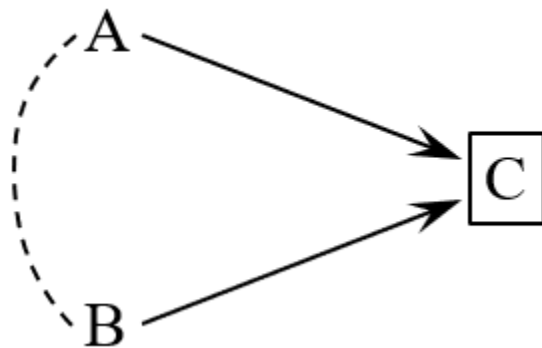


Pearl's Sprinkler Example

A: It rains

B: The sprinkler is on

C: The lawn is wet



Academic Tenure Example

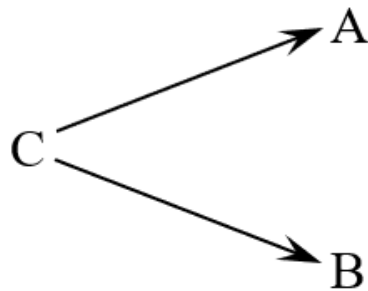
A: Productivity

B: Originality

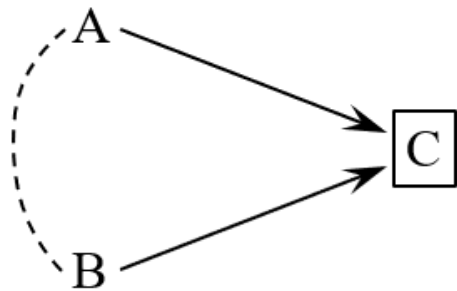
C: Tenure

Confounding vs. Selection Bias

Confounding bias and selection bias are different. Suppose that A is treatment and B is the outcome.



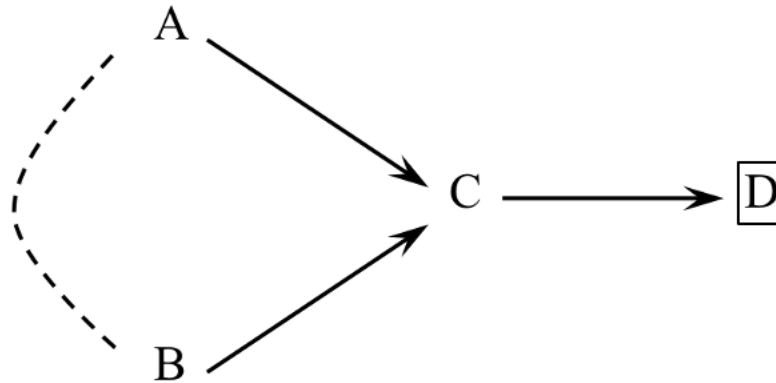
Confounding Bias: failure to condition on a common cause



Selection Bias: mistakenly conditioning on a common effect

Confounding and selection result from different analytic actions with respect to different causal structures, and they also have different solutions.

There's More: Conditioning on a Descendant of a Collider



Conditioning on the descendant, D , of a collider, C , results in the same problem as conditioning on C itself: It induces an association between the collider's parents.

D-separation: Formal Definition

We now have all we need. The rules for opening and closing paths of arbitrary length are summarized in Pearl's (1988) famous d-separation criterion. D-separation is the most important concept in DAG world.

Definition: A path is said to be **d-separated** (or **blocked** or **closed**) by a set of nodes Z iff the path contains

1. a chain $A \rightarrow M \rightarrow B$ or a fork $A \leftarrow M \rightarrow B$ such that M is in Z , or
2. an inverted fork $A \rightarrow M \leftarrow B$ and neither M nor any of its descendants is in Z .

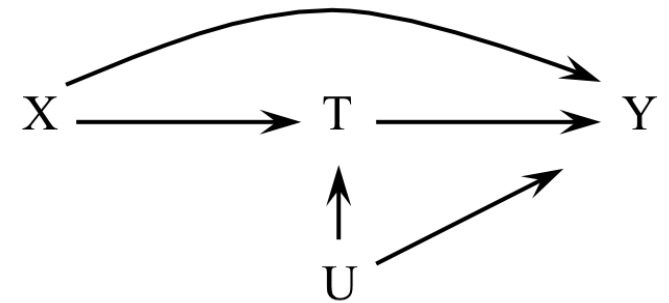
Definition: A path is **d-connected** (or **open** or **unblocked**) iff it is not d-separated.

FYI, the “d” in d-separation stands for “directional.”

D-separation and Statistical Independence

D-separation is a graphical property. Verma & Pearl (1988) relate d-separation to statistical associations. Specifically, they prove that, under mild conditions, in data generated according to the causal DAG where Z is conditioned on:²

1. A path **does not transmit association** (is closed or blocked) if it is d-separated by Z .
2. A path **may transmit association** (is open or unblocked) if it is d-connected.



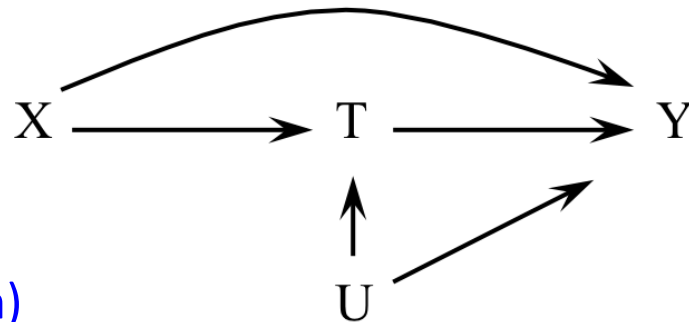
Implication A: **Two variables are statistically independent** if there's no open path between them.

Implication B: **Two variables may be statistically associated** if there is at least one open path between them.

² “Conditioning” here means “perfect stratification.” Sometimes, one can get away with interpreting “conditioning” more broadly as “introducing information into the analysis,” e.g., by regression control, coarse matching, or sample selection.

D-separation in Words

Whether a path between two variables is open or closed depends on whether the analyst does, or does not, condition on a collider or on a non-collider variable.



Blocking Rules (d-separation)

Rule 1: *Conditioning on any non-collider on a path blocks the path.*

Rule 2: *Not conditioning on at least one collider on a path (nor any of its descendants) blocks the path.*

Opening Rule (d-connection—the converse of d-separation):

Rule 3: *Not conditioning on any non-colliders and conditioning on all colliders (or on at least one descendant of each collider) on the path opens the path.*

D-separation Exercises

Suppose that the causal DAG represents the true DGP.

Are T and U associated marginally?

Yes, the path $U \rightarrow T$ is open.

Are X and Y associated marginally?

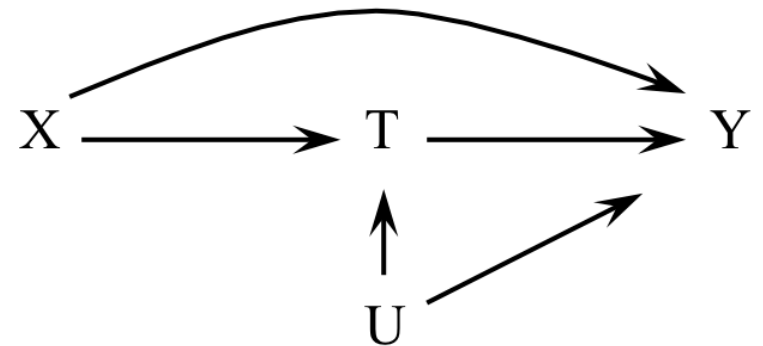
Yes, the two paths $X \rightarrow T \rightarrow Y$ and $X \rightarrow Y$ are open.

Are X and U associated marginally?

No, the path $X \rightarrow T \leftarrow U$ is blocked by the unconditioned collider T , and the paths $X \rightarrow Y \leftarrow U$ and $X \rightarrow T \rightarrow Y \leftarrow U$ are blocked by the unconditioned collider Y .

Are X and U associated after conditioning on T?

Yes, conditioning on T opens the path $X \rightarrow T \leftarrow U$ because T is a collider on this path.



Application: The Problem of Mediation Analysis

With these simple rules, we can understand some fundamental problems.

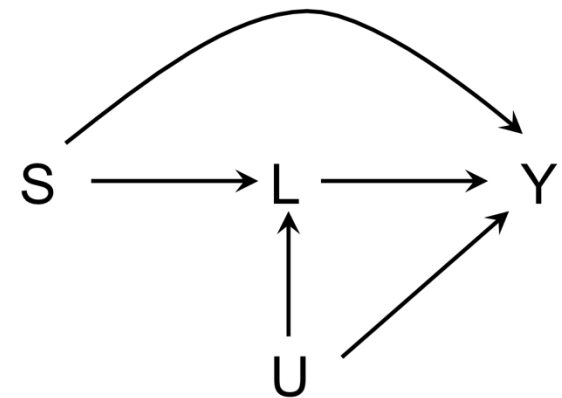
Suppose that the DAG captures the DGP. Here, S was randomized (no systematic input into S), but L was not randomized (systematic input into L from S and U).

Does the marginal association (i.e., no conditioning) between S and Y identify the total causal effect of S on Y ?

Yes: All causal paths ($S \rightarrow Y$ and $S \rightarrow L \rightarrow Y$) are open, and all non-causal paths ($S \rightarrow L \leftarrow U \rightarrow Y$) are closed!

Does the conditional association between S and Y given L identify the direct causal effect $S \rightarrow Y$?

No: Conditioning on the collider L opens the path $S \rightarrow L \leftarrow U \rightarrow Y$, because L is a collider on this path! Thus, S and Y would be associated given L even if there were no direct causal effect $S \rightarrow Y$.



Birthweight Paradox: Setup

Now let's review a real example. The "birthweight paradox" (Yerushalmy 1971) comprises two findings:

1. Maternal smoking in pregnancy is associated with both (a) low birthweight and (b) higher neonatal mortality on average.
2. But among low birthweight babies (<2500g), maternal smoking is associated with lower neonatal mortality.

The challenge is: How should we interpret #2?

- Is smoking in fact beneficial for low-birthweight babies (causation)?
- Or is it a methodological artifact (bias)?

VanderWeele (2014) hypothesizes that the birthweight paradox has exactly the structure that we analyzed before.

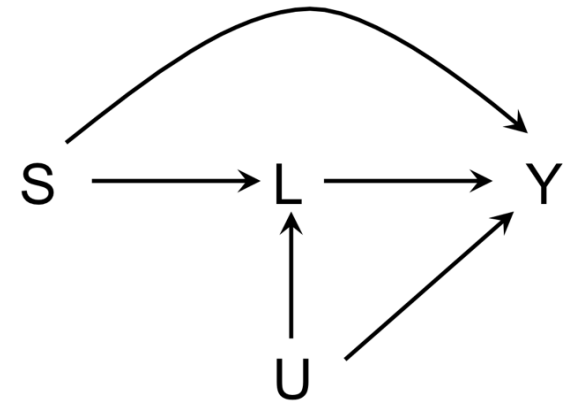
Selection Bias Due to Conditioning on a Mediator

S: Maternal smoking

L: Low birthweight

Y: Neonatal mortality

U: Common causes of low birthweight and mortality
(e.g., malnutrition, birth defects)



Stratification on L opens a non-causal path $S \rightarrow L \leftarrow U \rightarrow Y$.

If this path is negative, the association between S and Y given $L = 1$ could become negative even if the direct causal effect $S \rightarrow Y$ is positive. Research shows that this is empirically plausible (VanderWeele 2014).

Intuitively: If maternal smoking and malnutrition both cause low birthweight, then low-birthweight infants whose mother did not smoke likely are malnourished (and vice versa). If malnutrition is worse for mortality than maternal smoking, then low-birthweight infants of non-smoking mothers may have higher mortality.

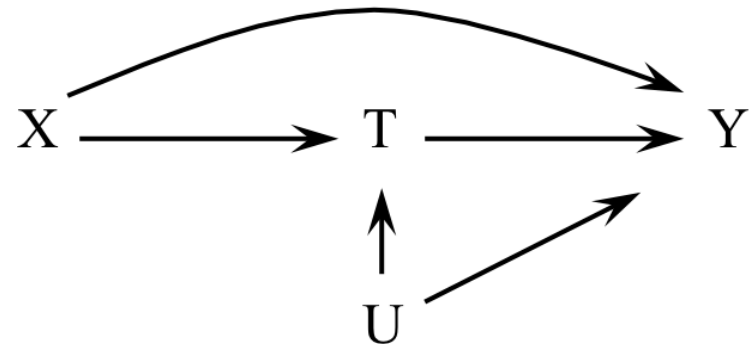
Testable Implications

Since we know that causal inference requires causal assumptions, it's important to test one's assumptions empirically as best possible.

D-separation is powerful: it derives all marginal and conditional independences implied by the DAG.

The implied independences involving only observed variables are testable.

The denser the DAG, the fewer testable implications there are. If U is latent, then this DAG contains no testable implications, because d-separation says that all observables (X, T, Y) may be marginally and conditionally associated.



If U were observed, then $X \perp U$ would be testable. If this independence is rejected by a suitable statistical test, then the DAG may be wrong and should be reconsidered.

Various software packages can enumerate testable implications of model, e.g., the dagitty package in R.

Causal Discovery

This begs the question whether you can “learn” the causal DAG from data alone. (This is the field of “causal discovery”.)

The answer is: No.

Even under ideal conditions (infinite sample size, no measurement error), one can never test all assumptions in a model.

Reason: we must assume that the DAG is “causal,” i.e., that it displays all common causes of all variables in the DAG. Since there is a (near) infinite number of potential common causes, one can never test out the absence of all possible common causes (Robins & Wasserman 1999).

Popular causal discovery algorithms therefore assume that all variables in the causal DAG are observed. That may be a tall assumption in an empirical application.

3. Inferring Causation from Association

Causal Identification

Next, we'll determine whether a total causal effect is identifiable by controlling (“adjusting”) for the right control variables, relative to the assumptions in a DAG.

Definition (intuitive): The average total causal effect (ATE) of T on Y is “causally identifiable” if it is possible to purge all non-causal association from the observed association between T and Y such that only the causal association remains.

Remember this intuition—it's very useful in practice.

The Adjustment Criterion

Ilyia Shpitser et al. (2010) have come up with a simple way to determine from a DAG whether the average total causal effect (ACE or ATE) of T on Y is identifiable by covariate adjustment.

Adjustment criterion (simplified): *The ATE of T on Y is “identified by adjustment” if one can adjust for (i.e., condition on, or control) a set of observed variables X such that*

- 1. X closes all non-causal paths between T and Y, and*
- 2. X does not close any causal paths between T and Y.*

Adjustment criterion (formal): A set of variables X (which may be empty) fulfills the adjustment criterion relative to the ATE of T on Y iff

1. X blocks all non-causal paths from T to Y, and
2. No element of X is on a causal path from T to Y or descends from a variable on a causal path from T to Y.

If X meets the adjustment criterion, then X is a “sufficient adjustment set.”

Summary: d-Separation, Independence, Identification

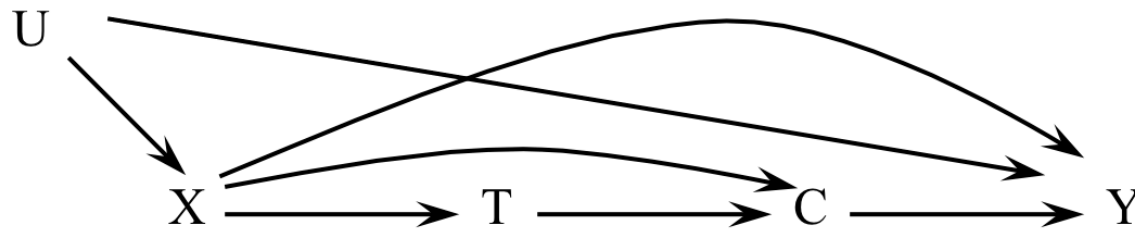
Here's how our three key concepts hang together so far:

d-separation (of a **path**): No association flows along the path.

Independence (of **variables**): d-separated along all paths.

Identification (of a **causal effect**): Treatment and outcome are d-separated along all noncausal paths and d-connected along all causal paths. (There are other notions of causal identification, which we haven't met yet.)

Example 1



Can you identify the total causal effect of T on Y by adjustment if U is unobserved?

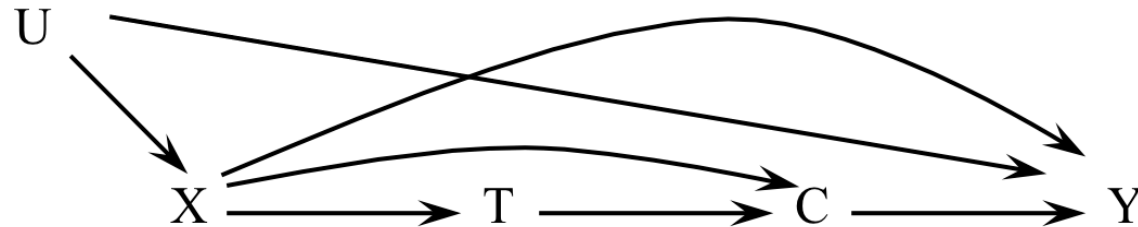
Yes. There are 5 non-causal paths from T to Y.

All 5 non-causal paths pass through X.

X is a non-collider on all paths.

Since conditioning on a non-collider blocks a path, conditioning on X blocks all non-causal paths from T to Y. And since conditioning on X does not block the causal path, conditioning on X identifies the total causal effect of T on Y.

Example 2



Can you identify the total causal effect of X on Y by adjustment if U is unobserved?

No. There is one non-causal path from X to Y: $X \leftarrow U \rightarrow Y$.

This non-causal path is open since U is a non-collider on the path.

Since U is unobserved, we cannot condition on U to block the path.

Hence, the observed association between X and Y is a mixture of (a) the total causal effect of X on Y (via three causal paths) and (b) the non-causal association via the open non-causal path $X \leftarrow U \rightarrow Y$.

Shortcuts

Application of the adjustment criterion can be a bit unwieldy in large DAGs. Helpful shortcuts exist. These include the

- Backdoor criterion
- Parent criterion
- Weak bow pattern criterion
- Strong bow pattern criterion

These are explained in Pearl (2009) and Elwert (2013).

For example, the parent criterion says that adjusting all parents of a treatment variable in a causal DAG is sufficient for identification by adjustment.

The parent criterion captures the popular intuition that one should “control for the treatment-assignment mechanism.”

Note that the parent criterion is sufficient, but not necessary. I.e., you don't generally have *to* control for all causes of treatment.

Adjustment and Conditional Ignorability

Here's a real gem: The adjustment criterion underlies causal claims with regression and matching estimators!

If X meets the adjustment criterion for the ATE of T on Y , then $\{Y(t)\} \perp T|X$.

This is a super useful result. You can now inspect the assumed DAG to determine whether adjustment for covariates X identifies the ATE of T on Y .

Non-parametric Estimation

If the adjustment criterion is met for the total causal effect of T on Y given X , then it implies a non-parametric estimator (Pearl 2009):

The distribution of the counterfactuals $Y(T)$, $P(Y(T))$, is given by

$$P(Y(T)) = \sum_X P(Y | T, X) P(X).$$

Then, for example, the ATE of switching a binary T from 0 to 1 on a binary Y with discrete covariates X is non-parametrically estimated as:

$$\begin{aligned} &Pr(Y(1) = 1) - Pr(Y(0) = 1) = \\ &\sum_X (Pr(Y = 1 | T = 1, X = x) - Pr(Y = 1 | T = 0, X = x))P(X = x). \end{aligned}$$

This estimator is sometimes known as the non-parametric g-formula. We also recognize it as the exact-matching estimator.

Unfortunately, analysts rarely have datasets large enough to implement it (because X tends to be high dimensional). In practice, analysts simplify this estimator by inserting functional-form assumptions. That's a big topic on its own.

Identification Beyond Adjustment: Do-calculus

In reality, the adjustment criterion often fails, i.e., one can't identify a causal effect simply by adjusting for observed covariates.

However, more powerful graphical identification criteria exist.

Pearl's (1995) do-calculus is the mother of them all. The do-calculus is "complete for the causal effects of interventions," i.e., it can detect whether or not the causal effect of a hypothetical external intervention on one or more treatment variables on one or more outcome variables is non-parametrically identifiable relative to a given DAG.

Various packages implement the do calculus, e.g., `causaleffects` in R.

4. Using DAGs to Shed Light on Things

A Selection of Applications

At this point, we've explored some of the main uses of DAGs.

1. Notating the qualitative causal structure of the DGP
2. Deriving testable implications of the model (d-separation)
3. Non-parametric identification analysis (adjustment criterion)

There's much more, of course. In the remainder of this workshop, I'll provide a subjective selection of applications, for which DAGs have proven useful. I make no pretense of completeness.

1. Selection bias
2. Interpreting regression coefficients
3. Mediation
4. Instrumental variables

4.1. Selection Bias

Endogenous Selection Bias

Conditioning on a collider—or a descendant of a collider—often leads to selection bias (e.g., Hernan et al., 2004).

Science is full of examples, where conditioning (as opposed to not conditioning) on a collider leads to bias, e.g.,

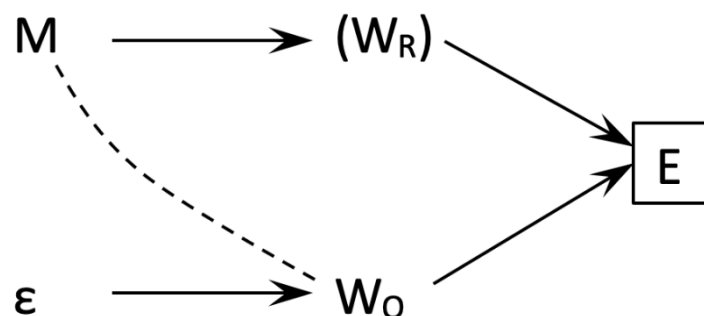
- Berkson's bias
- Ascertainment bias
- Induced confounding
- Dependent censoring
- Non-response bias
- Many more

Let's look at two real examples (simplified for clarity; more in Elwert and Winship 2014).

Motherhood Wage Penalty

Conditioning on an outcome, or a descendant of an outcome, ruins identification. Gronau (1974) and Heckman (1974) famously analyzed this in economics.

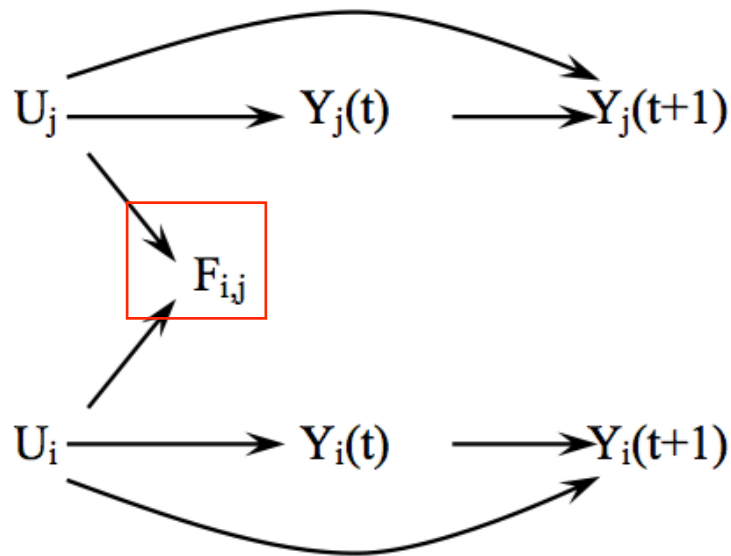
Consider the causal effect of motherhood, M , on employers' wage offers, W_O . In the standard labor model, motherhood increases women's reservation wage, W_R . If $W_O > W_R$, the woman accepts employment, $E = 1$. Suppose, for clarity, that M exerts no causal effect on W_O . The DAG captures this story.



Problem: Social scientists (to this day) often estimate the causal effect of M on W_O on samples of employed women, i.e., they condition on the collider E . This will induce a (non-causal) association between M and W_O , even if motherhood has no causal effect on wage offers. Hence, naïve regressions of W_O on M (and whatever other controls deemed important), will be biased.

Pre-treatment Collider Bias

Even controlling for pre-treatment variables can create bias—e.g., if the variable is a collider. This is the structure of “latent homophily” bias in social network analysis (Shalizi and Thomas 2011).



Y represents civic engagement for person i (Igor) and person j (Jane). There is no causal effect from Y_i on Y_j . Let's say that Igor and Jane are friends, $F_{ij} = 1$, because they are both into politics, U , which is unmeasured.

Now follow conventional practice in social network analysis and sample friendship dyads, i.e., select the sample on being friends, $F_{ij} = 1$.

This sample selection would open the non-causal path from exposure Y_i to outcome Y_j , $Y_i \leftarrow U_i \rightarrow F_{ij} \leftarrow U_j \rightarrow Y_j$ and make them associated even though there is no causal effect from Y_i on Y_j .

(O'Malley et al 2014 provide an instrumental-variables solution.)

4.2. Interpreting Regression Coefficients

Interpreting Regression Coefficients

Many empirical researchers run regressions to estimate causal effects. DAGs can help determine whether, and which, the regression coefficients identify causal effects (Keele et al. 2019).

The process goes like this:

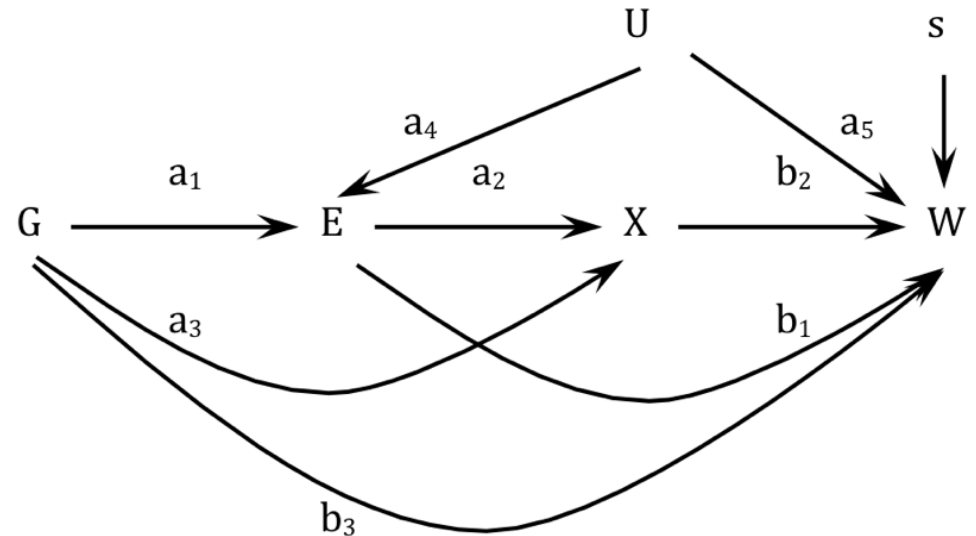
1. Commit to a DAG to represent the assumed DGP
2. Make additional functional form assumptions
3. Specify a regression
4. Determine which paths contribute to each regression coefficient.

Example: Wage Determinants

Suppose that the DGP by which gender, G , education, E , experience, X , and ability, U , determine log wages, W has the following structure:

Suppose further that this DAG represents a linear and homogenous DGP (crazy, I know), so that the path coefficients on each arrow represent causal effects in the regression metric.

All variables except U are observed, and s is the idiosyncratic error term on W .



Now we estimate a conventional wage regression,

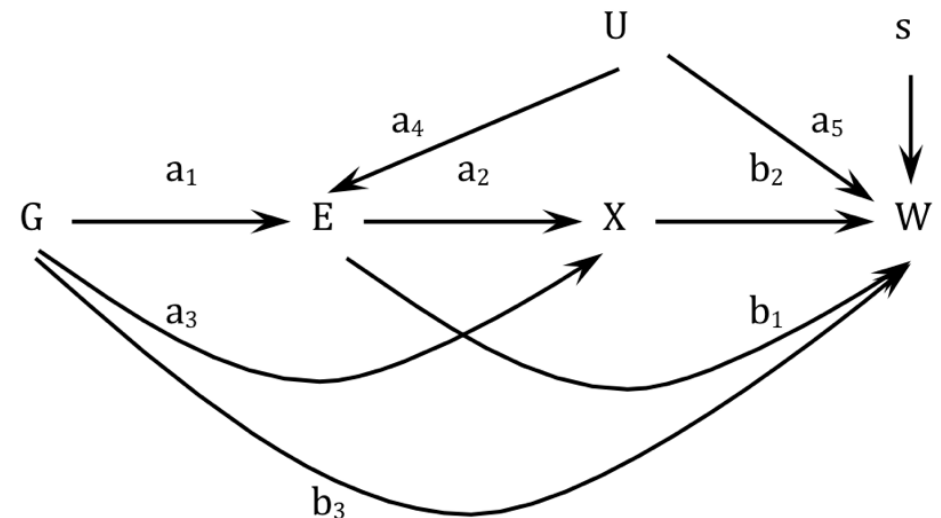
$$\ln(Wage) = Experience \beta_1 + Education \beta_2 + Gender \beta_3 + e$$

Note that the regression error term, $e = s + U$, contains U .

Finally, use d-separation to figure out which paths contribute to each regression coefficient.

$$\ln(\text{Wage}) = \text{Experience} \beta_1 + \text{Education} \beta_2 + \text{Gender} \beta_3 + e$$

1. β_1 captures the association flowing along the open path between X and W (conditional on E and G): $X \rightarrow W$, which happens to be the only causal path from X to W . Hence, β_1 identifies the total causal effect of X on W under this model.



2. β_2 captures the association flowing along the open paths between E and W (conditional on X and G): the direct causal effect $E \rightarrow W$, and the confounding path $E \leftarrow U \rightarrow W$. Since the latter path is non-causal, β_2 does not identify a causal effect under this model.

3. β_3 captures the association flowing along the open paths from between G and W (conditional on E and X): the direct causal effect $G \rightarrow W$ and the non-causal path $G \rightarrow E \leftarrow U \rightarrow W$, which is open because the regression controls for the collider E (selection bias). Hence, β_3 does not identify a causal effect under this model.

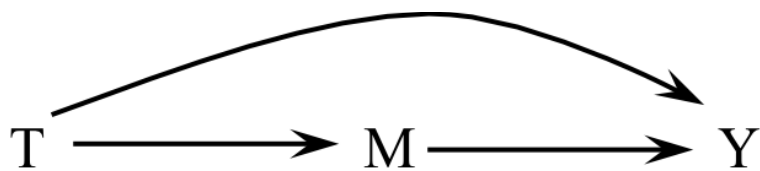
4.3. Causal Mediation Analysis

Causal Mediation Analysis

Causal mediation analysis remains a frontier of causal inference. DAGs have been instrumental in recent developments (Pearl 2001, VanderWeele 2015).

Earlier, in this workshop, we already met the “fundamental difficulty” (Imai et al. 2010) of causal mediation analysis—unobserved mediator-outcome confounding.

Methodologists have distinguished several subtle different notions of mediation effects (Nguyen et al. 2020), including:



Controlled direct effects (CDE): How would intervention on T affect Y if M was set to some externally chosen value, $M = m$?

Natural direct effect (NDE): How would intervention on T affect Y if M was set to the value that it would have if T was set to its baseline value, $M = M(t = 0)$?

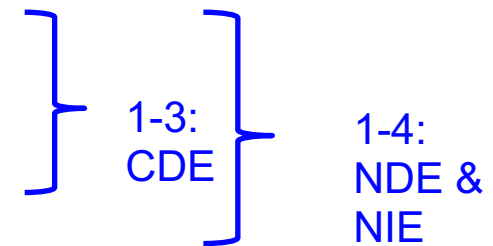
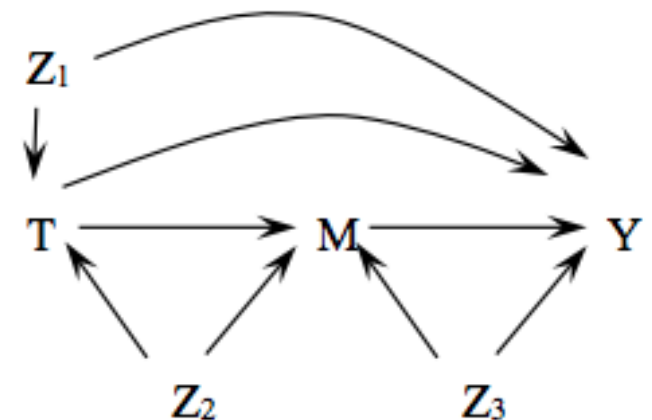
Natural indirect effect (NIE): How would Y change if M changed in the amount that it would change if T was intervened upon ($M(t = 1)$ vs $M(t = 0)$)?

Adjustment Criteria for CDE, NDE, and NIE

Most analysts use some form of regression modeling to estimate CDE, NDE, or NIE. With DAGs, it's easy to express the causal identification requirements for regression- and other adjustment-based estimators.

Adjustment Criteria for Mediation: Causal mediation effects of T on Y relative to the (set of) mediator(s) M are identifiable by adjusting for a set of variables Z if

1. Z does not include descendants of T ,
2. Z and M block all backdoor paths from T to Y ,
3. Z and T block all backdoor paths from M to Y ,
4. Z blocks all backdoor paths from T to M .



Note: a backdoor path from V is a path that begins with an arrow into $\rightarrow V$.

E.g., in this DAG, the model would need to adjust for Z_1 , Z_2 , and Z_3 in order to identify the NDE and NIE. By contrast, it would only have to adjust for Z_1 and Z_3 to estimate the NDE. See VanderWeele (2015) for a survey of estimation approaches and more general identification criteria.

4.4. Instrumental Variables

Instrumental Variables Analysis

In many real applications, identification by adjustment fails, because treatment, T , and outcome, Y , are confounded by an unobserved variable, i.e., there is a non-causal path $T \leftarrow U \rightarrow Y$, which cannot be closed by adjustment.

Even though adjustment fails, however, there sometimes exist so-called “instrumental variables” (IV) that permit the identification of the causal effect $T \rightarrow Y$ under certain functional constraints (e.g., linearity).

DAGs are useful for detecting IVs in complex DGPs.

DAGs are also useful for illuminating thorny problems in IV analysis, and are now widely used for this purpose, e.g., in epidemiology.

Graphical Instrumental Variables Criterion

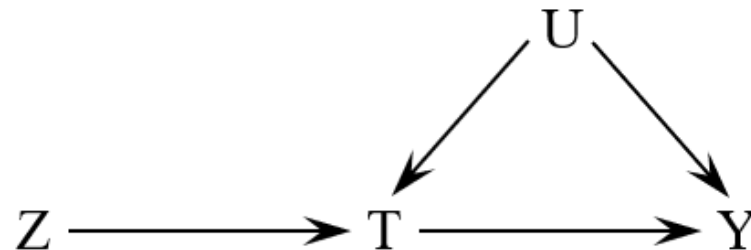
There are several subtly different definitions of instrumental variables (Swanson et al. 2018). Here is a basic purely graphical definition (Pearl 2001):

Definition: A variable, Z , is called an instrumental variable for the causal effect of T on Y , if, conditional on covariates X (which may be empty),

A1: There is at least one open path from Z to T conditional on X ,

A2: X does not contain descendants of Y ,

A3: There is no open path from Z to Y conditional on X , other than those paths that terminate in a causal path from T to Y .



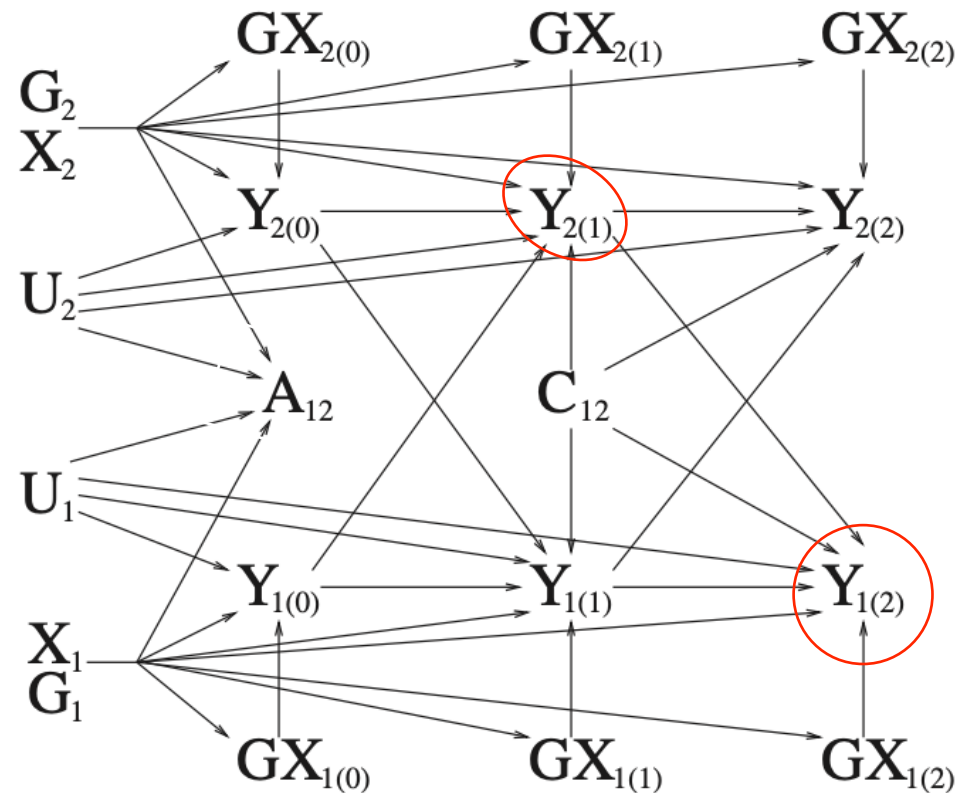
This is the canonic IV DAG, where Z is a randomized IV for the effect of T on Y . Brito and Pearl (2002) provide much more general graphical IV criteria.

Example: Genes as IVs for Peer Effects

O'Malley et al. (2014) used various graphical IV criteria to detect IVs in a fairly complex DAG for peer effects of one person's body mass index (BMI), Y_2 , on another person's BMI, Y_1 , in a longitudinal social network.

The causal effects $Y_{2(t)} \rightarrow Y_{1(t+1)}$ cannot be identified by adjustment, because the data selected on dyads, $A_{12} = 1$, and the factors determining social ties, U_j , and shared contextual factors, C_{12} , are latent.

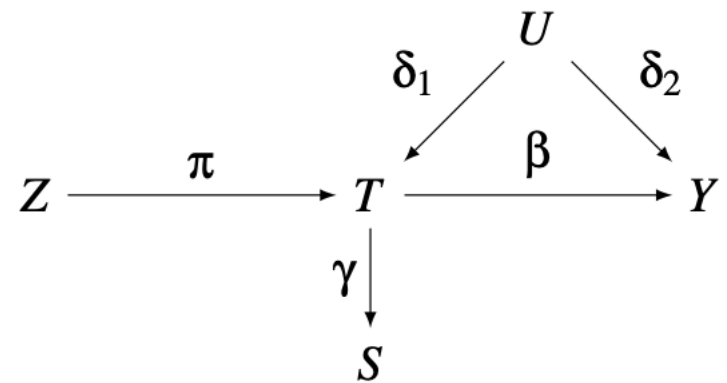
Regardless, it can be shown that person 2's age-varying gene expressing in period t , $GX_{2(t)}$, is a valid IV for the effect of her BMI, $Y_{2(t)}$, on her peer's subsequent BMI, $Y_{1(t+1)}$, conditional on her own genes, G_2 , covariates, X_2 , and $A_{12} = 1$. This result holds under this and several more general DAGs.



DAGs for Instrumental Variables under Selection

The graphical logic d-separation is useful for detecting and understanding overlooked problems in IV analysis. For example, selecting the sample as a function of treatment, $S = f(T)$ necessarily invalidates Z as an IV:

Since S is a descendant of the collider T on the path $Z \rightarrow T \leftarrow U \rightarrow Y$, Z is associated with Y (“not excluded”) via a path that does not terminate in $T \rightarrow Y$.



Graphical intuition also helps with quantifying biases. E.g., Elwert and Segarra (2020) investigate several linear IV models with selection on S via various selection mechanisms (adjustment or truncation) and show the biases map precisely onto the non-causal paths that are opened or closed by selection.

Quantitative analysis pays off. For this DAG, they show that the OLS and IV estimators with selection, though both biased, bound the true causal effect $T \rightarrow Y$.

Summary and Outlook

DAGs are a popular tool for causal inference. Their main uses are:

1. Notating the causal structure of the DGP
2. Deriving testable implications of a model (d-separation)
3. Non-parametric causal identification analysis (do-calculus)

DAGs have contributed to methodological developments in many applied fields, e.g.,

1. Understanding and recovering from selection bias
2. Causal mediation
3. Instrumental variables
4. Transportability/Generalizability
5. Missing data
6. Causal discovery

Maathuis et al. (2018) are a recent comprehensive survey of graphical models.

Online Courses and Textbooks

- Courses on DAGs:
 - Via Statistical Horizons: Various Courses on DAGs and Causal Inference (<https://statisticalhorizons.com/our-instructors/felix-elwert>)
 - Via edX: Draw Your Assumptions Before Your Conclusions (<https://www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your>)
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
- Pearl, J., and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

References

- Brito, C. (2010). Instrumental sets. In Dechter, R., Geffner, H., and Halpern, J., editors, *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, chapter 17, 295-307. College Publications.
- Brito, C. and Pearl, J. (2002). Generalized instrumental variables. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, Canada, 85-93.
- Elwert, F. (2013). *Graphical Causal Models*, chapter 13, 245-273. Springer.
- Elwert, F. and Segarra, E. (Forthcoming). Instrumental Variables with Treatment-Induced Selection: Exact Bias Results. In Dechter, R., Geffner, H., and Halpern, J., editors, *Probabilistic and Causal Inference: The Works of Judea Pearl*. ACM Books.
- Elwert, F. and Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40(1), 31-53.
- Gronau, R. (1974). Wage comparisons--a selectivity bias. *Journal of Political Economy*, 82(6), 1119-1143.
- Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, 42(4), 679-694.

Hernan, M. A., Hernandez-Diaz, S., and Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology*, 15(5), 615-625.

Hernan, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15(4), 309-334.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, first edition.

Keele, L., Stevenson, R. T., and Elwert, F. (2019). The causal interpretation of estimated associations in regression models. *Political Science Research and Methods*, 8(1), 1-13.

Maathuis, M. H., Drton, M., Lauritzen, S., and Wainwright, M., editors (2018). *Handbook of Graphical Models*. CRC Press, first edition.

Morgan, S. and Winship, C. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.

Nguyen, T. Q., Schmid, I., and Stuart, E. A. (2020). Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. *Psychological Methods*.

O'Malley, A. J., Elwert, F., Rosenquist, J. N., Zaslavsky, A. M., and Christakis, N. A. (2014). Estimating peer effects in longitudinal dyadic data using instrumental variables. *Biometrics*, 70(3), 506-515.

Pearl, J. (1985). Bayesian networks a model of self-activated memory for evidential reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*, University of California, Irvine, CA, USA.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc.

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

Pearl, J. (2001). Parameter identification: A new perspective. Technical Report R-276, UCLA Cognitive Systems Laboratory.

Pearl, J. (2009). *Causality*. Cambridge University Press, second edition.

Pearl, J. (2013). Linear models: A useful microscope for causal analysis. *Journal of Causal Inference*, 1(1), 155-170.

Richardson, T. S. and Robins, J. M. (2013). Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. CSSS Technical Report No.128.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2), 143-155.

Robins, J. M. and Wasserman, L. (1999). On the Impossibility of Inferring Causation from Association without Background Knowledge. In Glymour, P., and Cooper, G., editors, *Computation, Causation, and Discovery*, 305-321. The MIT Press.

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

Shalizi, C. R. and Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2), 211-239.

Tian, J. and Shpitser, I. (2010). *On Identifying Causal Effects*, chapter 29, 415-444. College Publications.

VanderWeele, T. J. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.

Verma, T. and Pearl, J. (1988). Causal networks: Semantics and expressiveness. In *Proceedings of the Fourth Annual Conference on Uncertainty in Artificial Intelligence*, 69-78. North-Holland Publishing Co.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, XX(7), 557-585.

Yerushalmy, J. (1971). The relationship of parents' cigarette smoking to outcome of pregnancy implications as to the problem of inferring causation from observed associations. *American Journal of Epidemiology*, 93(6), 443-456.