Fall 2021 Causal Inference reading group

Time: Thursday 10-11am MT
Zoom link: https://uwyo.zoom.us/j/3076453097
Google drive folder

Link to free online version of Causal Inference: The Mixtape
https://mixtape.scunning.com/

Schedule

Week	Meeting date	Reading	Moderator
1	9/16	Ch 1 - Introduction, Ch 3 - Directed Acyclic Graphs (DAGs)	Andrew
2	9/23	Ch 4 - potential outcomes (PO)	
3	9/30	Causal inference in experiments - Kimmel et al. 2021 TREE Kendall 2015 book chapter (suggested)	Including Paul Ferraro
4	10/7	Causal inference in observational studies - Larsen et al. 2019 MEE	
5	10/14	Ch 5 - Matching and Subclassification	
6	10/21	Brandt et al. 2015 - matching methods to assess effectiveness of protected areas	with Katherine Siegel
7	10/28	Angrist & Pischke - Mostly Harmless Econometrics Ch5 - Fixed Effects, DiD, and Panel Data	
8	11/4	Dudney et al. 2021 - panel data to assess climate effects on disease range shifts	Joan Dudney
9	11/11	Laura's paper using panel data to estimate biodiversity effect on productivity - **IN REVIEW, DO NOT SHARE** Manuscript, SI, RMarkdown tutorial, data	Laura Dee
10	11/18	Kendall 2015 - Instrumental variables	
11	12/2	MacDonald et al. 2019 - IV to estimate effect of forest fragmentation on Lyme disease incidence	
12	12/9	IMF Working Paper - regression discontinuity analysis of mask mandate effectiveness Optional background reading: Mixtape Ch 6	

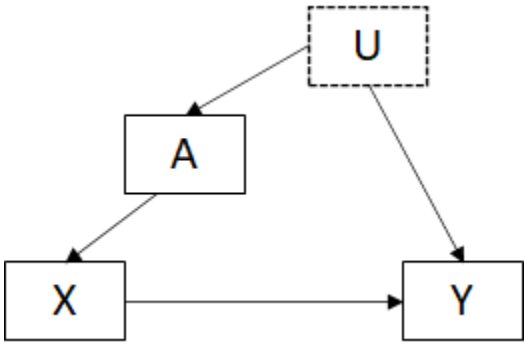
Youtube channel -- https://jhsphcausalinference.weebly.com CAUSAL INFERENCE WORKING GROUP, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Week 10: Kendall 2015 - Instrumental variables

Add topics/questions for discussion below

Andrew: What's the difference between consistency and bias?

Andrew: "when performing multiple regression (more than one predictor variable), a single endogenous variable can cause the estimated coefficients of the other predictor variables to be inconsistent as well". I was thinking about this in terms of the backdoor criterion. Consider this DAG where X is the predictor of interest (we want to measure its causal effect), Y is the response variable, and U is an unmeasured variable. Conditioning on A closes the one backdoor between X and Y, which I thought would let us estimate the causal effect of X on Y. But if I understand correctly, A is endogenous (correlated with U, which is part of the error term), so based on the statement above, this causes the estimated effect of X on Y to be inconsistent.



Week 9: Laura's paper (in review) and RMarkdown tutorial

Add topics/questions for discussion below

Andrew: Could you have used natural (rather than experimentally manipulated) species richness in the area around a plot as an instrumental variable? Or would there still potentially be a problem of reverse causality, i.e., plot-level productivity influences species richness surrounding the plot? Or some other spurious pathway I'm not thinking of?

Andrew: On the bottom of pg SI-5 it says: "The use of random effects requires the assumption that the random effect is uncorrelated with all of the covariates in the model." I've been confused by statements like this. Does this really mean something like "The use of random effects *to control for unobserved group-level differences when trying to estimate a within-group causal effect* requires the assumption…"

Andrew: On SI-13 it says. “Note that delta_ps is not part of the error term, as it would be in mixed (multi-level) models. Rather, it is a parameter to be estimated.” Statements like this have also caused confusion for me, because when I specify a random effects model, the group-level random effects are treated as parameters to be estimated, not part of the error term (Gelman and Hill section 12.5 talks about 5 ways to write the same random effects model). I would specify a random effects version of Laura’s model like this:

$$\ln LiveMass_{pst} = \beta \ln Richness_{pst} + \delta_{ps} + \mu_{st} + \varepsilon_{pst}$$

δ_{ps} ~ Normal(μ_{δ}, σ_{δ}^2)
μ_{st} ~ Normal(μ_{μ}, σ_{μ}^2)

The only difference is that the random effects are assumed to come from a distribution that is itself estimated from the data (whereas fixed effects can be thought of as coming from a distribution with infinite variance). The random effects are parameters to be estimated, and they are not part of the error term in this specification.

Week 8: Dudney et al. 2021

Add topics/questions for discussion below

Andrew: Confused by: “To control for temporal serial correlation in the error terms within plots, standard errors were clustered by the plot”. Doesn’t the plot fixed effect take care of this?

Week 7: Fixed effects, DID, and panel data

Add topics/questions for discussion below

Andrew: I was still having trouble wrapping my head around the difference between fixed effects and random effects models and why random effects models have problems with bias. I found this paper and this one that clarified it for me somewhat, and also argue that using a different specification of a random effects model fixes the bias issue and has advantages over a fixed effects model. Maybe we could discuss one of these papers later in the semester if people are interested?

Alex: unfortunately I will need to sit out this week. Have fun. I found the reading fairly concrete this week. Main gap for me was more distilled guidance on how to evaluate the choice of different models or approaches.

More papers on fixed vs. random effects:
https://link.springer.com/article/10.1007/s11135-018-0802-x
https://link.springer.com/article/10.1007/s11135-017-0593-5

Week 6: Brandt et al. 2015 - Matching methods to assess effectiveness of protected areas

Add topics/questions for discussion below

Alex:

- I looked up 'Shangrila' from Control Variables. It is a city named after the fictional city with that name, to promote tourism (see https://en.wikipedia.org/wiki/Shangri-La_City).
- I thought the paper was remarkably clear and well-reasoned.
- I think it would be good to have some measures of the quality of the matches that are available in a data set.
- "We test for the quality of our matches by calculating the amount of bias reduced between matched and unmatched pairs". How do they quantify the bias? For example, here is one table entry from the supplement:

Variable		Treated	Control	% bias	% bias reduction	T-stat	p-value
Distance to shangrila	Unmatched	58915	64064	-22.5		-11.49	0
	Matched	58915	62178	-14.2	36.6	-14.52	0

Andrew: I looked back at the Larsen et al. paper for a review of the DID model. It looks like what Brandt et al. refer to as a DID model:

$$FC_{it} = \delta(Treat_{it}) + X'_{it}\beta + \theta_t + \gamma_i + \varepsilon_{it}$$

Is equivalent to what Larsen et al. refer to as a within-estimator model (where fragmentation is the treatment):

$$Y_{it} = \alpha + \beta\,\text{Fragmentation}_{it} + c_i + \gamma_t + \varepsilon_{it},$$

This is what Larsen et al. call a difference-in-difference model:

$$Y_{igt} = \alpha + \delta_1 \text{treat}_g + \delta_2 \text{after}_t + \beta\left(\text{treat}_g \times \text{after}_t\right) + \varepsilon_{igt}$$

I'm trying to wrap my head around how these are different (or if they're equivalent).

Week 5: Mixtape Ch 5 - Matching and Subclassification

Add topics/questions for discussion below

Andrew: conditional independence assumption… same thing as ignorability and closing backdoor paths?

Andrew: how would matching methods apply to a continuous treatment variable? Could you estimate a propensity score for a continuous treatment variable?

Alex: I won't be able to make it this week as I am a bit under the weather. Sorry I will miss hearing you all fill in some of the gaps from the chapter. In particular, I appreciated the Titanic example, but would have liked to have a discussion of when conditioning on gender and age would incorrectly remove evidence of a true effect of 'first class' on mortality. If I understood correctly, the nominally biased, uncorrected estimate of the effect of 'first class' could completely disappear (and be falsely low) if the weighting from the covariates were aligned to make the 'first class' effect entirely nested within the confounder(s). Maybe I missed it, but I would have liked a discussion of false negatives (or underestimates of true effects) that we'd typically have in stats. More generally, I found that again reviewers would have been helpful to encourage the author to write some connections (at the outset of the chapter and throughout) to work other than foundational work in the 1900s, including contemporaries and better ties to earlier chapters and arguments in the book.

Week 4: Larsen et al. 2019 - Causal analysis in control–impact ecological studies with observational data

Add topics/questions for discussion below

Andrew: I thought this is an interesting way to think about when you might want to use the PO or causal graph framework depending on the goals of your study (from p925): "the PO framework… places an emphasis on isolating the effect of usually a single explanatory variable of interest (i.e. treatment variable) on a single outcome rather than on disentangling complex relationships within a network."

BACI = Before-After Control Impact = diff-in-diff. See BACI example. https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1939815 for example as a classic reference

Andrew: Does anyone have much experience with Granger causality? Is this something we want to explore in this reading group?

Andrew: Fixed and random effects...
Alex: You have probably seen this: "Why I don't use the term "fixed and random effects"" — https://statmodeling.stat.columbia.edu/2005/01/25/why_i_dont_use/

From Laura: Random effects assumption for unbiased estimate --

This is from Econometric Analysis of Cross Section and Panel Data, by Wooldridge:

The basic **unobserved effects model (UEM)** can be written, for a randomly drawn cross section observation i, as

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \qquad t = 1, 2, \ldots, T, \tag{10.11}$$

where $\mathbf{x}_{it}$ is $1 \times K$ and can contain observable variables that change across t but not i, variables that change across i but not t, and variables that change across i and t. In addition to unobserved effect, there are many other names given to c_i in applications: **unobserved component, latent variable,** and **unobserved heterogeneity** are common. If i indexes individuals, then c_i is sometimes called an **individual effect** or **individual heterogeneity**; analogous terms apply to families, firms, cities, and other cross-sectional units. The u_{it} are called the **idiosyncratic errors** or **idiosyncratic disturbances** because these change across t as well as across i.

(...)

In modern econometric parlance, a **random effects framework** is synonymous with zero correlation between the observed explanatory variables and the unobserved effect: $\text{Cov}(\mathbf{x}_{it}, c_i) = \mathbf{0}$, $t = 1, 2, \ldots, T$. (Actually, a stronger conditional mean independence assumption, $\text{E}(c_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = \text{E}(c_i)$, will be needed to fully justify statistical inference; more on this subject in Section 10.4.) In applied papers, when c_i is referred to as, say, an "individual random effect," then c_i is probably being assumed to be uncorrelated with the $\mathbf{x}_{it}$.

In most microeconometric applications, a **fixed effects framework** does not actually mean that c_i is being treated as nonrandom; rather, it means that one is allowing for arbitrary dependence between the unobserved effect c_i and the observed explanatory variables $\mathbf{x}_{it}$. So, if c_i is called an "individual fixed effect" or a "firm fixed effect," then, for practical purposes, this terminology means that c_i is allowed to be correlated arbitrarily with $\mathbf{x}_{it}$.

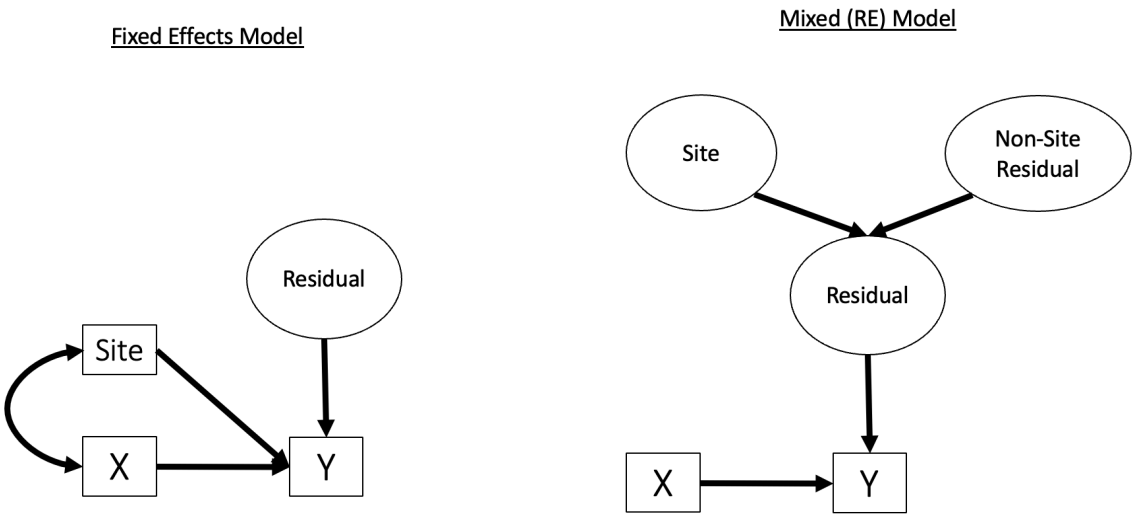
Share Cite Improve this answer Follow edited Jul 31 '13 at 16:01 answered Jul 31 '13 at 15:52

LD: IV papers -

- Example -- https://www.pnas.org/content/116/44/22212/tab-article-info
- Great overview by Imbens - https://www.jstor.org/stable/43288511?seq=1#metadata_info_tab_contents

Joan - within estimator example - I have to go and moderate a session but if anyone wants tor read an example of a fixed effect panel model approach using ecological data, I just published this paper https://www.nature.com/articles/s41467-021-25182-6

From Jarrett:



Week 3: Causal inference in experiments - Kimmel et al. 2021 TREE

Add topics/questions for discussion below

Andrew: From 1st paragraph on p4: **"**To control them statistically, one could try to measure the source of the violation (e.g., soil aeration) and control for it directly by including it as a covariate in a regression, similar to what is done in observational studies"
Wouldn't this be conditioning on a post-treatment variable?
Andrew: Do most excludability issues go away if you simply define the treatment as the manipulation rather than the thing you intended to manipulate? For example, in the drought experiment, you could define the treatment as shelter vs. no shelter (or shelter that doesn't let rain through vs. shelter that does let rain through) rather than drought vs. no drought.

Andrew: Could anyone provide an ecological example of a placebo outcome? Does it need to be something related to the actual outcome of interest?

Things that came up during the meeting:
https://jhsphcausalinference.weebly.com CAUSAL INFERENCE WORKING GROUP, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Causal inference in coupled human and natural systems: https://www.pnas.org/content/116/12/5311.short (see supplement for information about how to deal with interference). Partial identification

Interference -

- Manski - partial identification and bounding causal effects.
- Spatial lag model (vs spatial autocorrelation)
- Test if a control plot surrounded by treat, vs control plot surrounded by controls has the same average Y in expectation

That's a randomized saturation design in that article--
https://www.mitpressjournals.org/doi/pdf/10.1162/rest_a_00716

JB Seems like there is also an R package that works with this problem - "`inferference: Methods for Causal Inference with Interference`'
https://arxiv.org/pdf/2001.05444.pdf and
https://cran.r-project.org/web/packages/inferference/index.html - having not dealt with this problem before, this looks quite interesting

Week 2: Ch 4 - Potential Outcomes

Add topics/questions for discussion below

Check out Zach's awesome paper using DAGs

Next week - special guest Paul Ferraro

Plan for semester - mixtape book vs. papers; designs of interest; paper recommendations

Key concepts from Ch 4: potential outcomes, average treatment effect (ATE), average treatment effect for the treatment group (ATT), average treatment effect for the control/untreated group (ATU), simple difference in means (SDO), selection bias, heterogeneous treatment effect bias, independence assumption, stable unit treatment value assumption (SUTVA), randomization inference

Andrew: How do endogeneity, backdoor paths, independence assumption, ignorability (concept I've come across in Bayesian context) relate? Do they all refer to the same thing?

Alex: This is a bit of cleanup for me for concepts from last week and stems from my watching Richard McElreath's recent 3 hour causal inference video. DAGs are drawings that represent our understanding of science. The triads that exist within them have implications for how to do the statistical modeling (or where to apply randomization in the design of experiments), so it is good to be able to recognize the three types of triads that might exist in our graphs:

1. pipe : x → y → z
    a. x and z are dependent, and are connected in a causal path through y; knowing y is sufficient to predict z, but we might be interested in the strength of associations in the path

2. Fork: x ← y → z
    a. x and z are dependent, because both are genuinely caused by y; conditioning x on y correctly shows us that z is not caused by x; the association that exists between x and z is driven by y.
3. Collider: x → y ←z
    a. x and z are genuinely independent; conditioning x on y induces an association between x and z (i.e., if conditioned on y, if you tell me x, I can tell you something about z, even though it is independent of x otherwise.
    b. Don't condition on these.

Alex: R.A. Fisher's first name is Ronald, not Roland as evidently entered in the author's citation manager. I stumbled on other more important typos in the text.

Collider vs post-treatment bias or are they the same? The post treatment Bias would occur if you are estimating the treatment effect and conditioning on realized diversity - https://cpb-us-e1.wpmucdn.com/sites.dartmouth.edu/dist/5/2293/files/2021/03/post-treatment-bias.pdf

Yup - agreed - McElreath's Causal DAG and Haunted Terror chapter talks about this really well. https://bookdown.org/content/4857/the-haunted-dag-the-causal-terror.html

Week 1: Ch 1 - Introduction, Ch 3 - Directed Acyclic Graphs

Add topics/questions for discussion below

From Laura Dee - Other Resources- To go along with Ch. 3 on DAGS. I have always found Pearl's do-calculus stuff to be hard to understand; this blog post and associated book chapter looks very useful for breaking it down in plain english. https://www.andrewheiss.com/blog/2021/09/07/do-calculus-backdoors/

Other reading on DAGs/Pearl's causality framework - The Book of Why

1.3 (Andrew): What does endogenous mean?

Backdoor not being closed == endogeneity

Useful papers clarifying endogeneity more in the ecological + causal inference context -

- https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13600
- https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13190
- Kendall 2015 book chapter

Useful example from Aaron -- **Endogeneity and exogeneity are relative to a specification of a model** An example??: If you are predicting crop output in a farming system from a variety of factors, rainfall is exogenous, it is independent of other factors that may impact crop output. If Pest infestation and fertilizer use are also part of your model, pest infestation may be endogenous because it is may also influenced by fertilizer.

Here's another paper from grace and irvine that maybe translates some of this terminology to ecology (haven't read it, just skimming in a quick search)
https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.2962

3.1.5 (Andrew): In the gender discrimination example, it says there is no strategy to satisfy the backdoor criterion. But it appears there are no open backdoors to begin with since D->O->Y is not a back door and D->O<-A->Y contains a collider. Am I missing something?
LD: I understood this as … only the relationship between O and Y were considered by Google, missing the effect of discrimination because D → O and O is on the causal path to Y (earnings). So, by controlling for O, we get a new bias **also known as a bad control (I think) if we are trying to estimate the effect of D on Y, we shouldn't include O inthe model. I really don't totally understand how A plays into the example though, agree A is unobservable and therefore not controlled for. This relates to my question below too. I think it's O that is supposed to be the collider on D and Y, and A on Y and O so only A is OK in this example, but controlling for O leads to a biased estimate of D on Y.

(Laura) Can we discuss examples of colliders in ecology? I also couldn't tell if the suggestion is to leave the collider as is, or to include it in the analysis to condition on it? I have been taught that controlling for a collider is a "bad control" and creates new sources of bias. I think that is what is being said but could someone confirm their readings of this? I have heard the term "bad controls" - it's mentioned briefly, but I think it's a more intuitive one here for what to do.

Laura -- I don't think this very important point is widely appreciate in the use of SEM and path analysis in ecology - A DAG is meant to describe all causal relationships relevant to the effect of on . What makes the DAG distinctive is both the explicit commitment to a causal effect pathway and the complete commitment to the lack of a causal pathway represented by missing arrows. In other words, a DAG will contain both arrows connecting variables and choices to exclude arrows. And the lack of an arrow necessarily means that you think there is no such relationship in the data—this is one of the strongest beliefs you can hold. A complete DAG will have all direct causal effects among the variables in the graph as well as all common causes of any pair of variables in the graph.
I think omitted variable bias is vastly under-estimated in ecology -- and just because we haven't included an unobserved variable in a DAG/SEM, doesn't mean it isn't confounding. That is an assumption that a researcher is making.

Andrew - If the goal is to estimate a causal effect, should you only condition on covariates needed to satisfy the backdoor criterion? Is there benefit/harm in conditioning on variables

unrelated to backdoors, e.g. variables that only affect the outcome or only affect the causal variable?

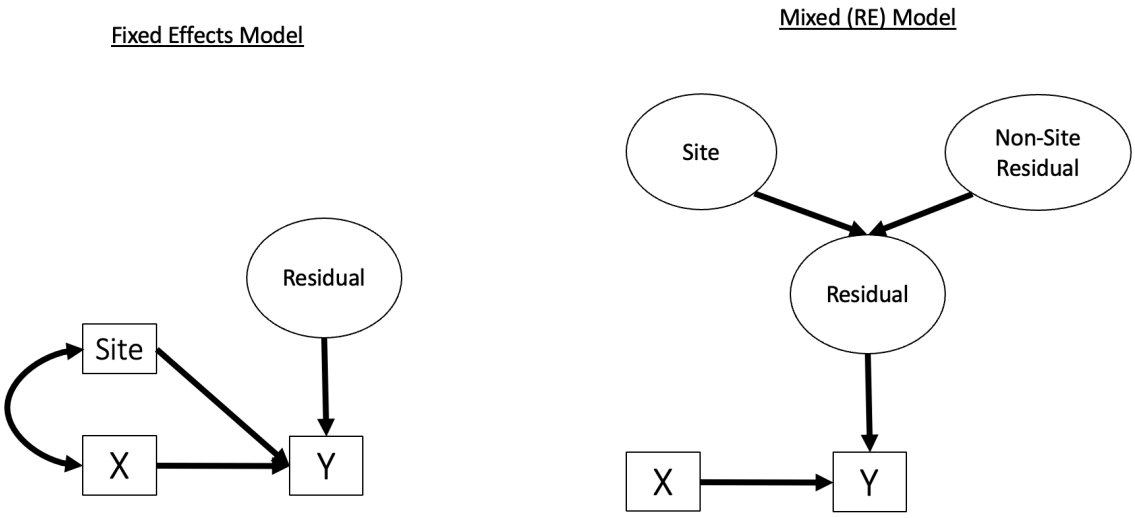
Larsen et al MEE - https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13190

Aaron (Farming Example): http://www-personal.umd.umich.edu/~delittle/Encyclopedia%20entries/Endogenous%20variable.htm

This is super helpful for DAGs + Simpson's paradox by Pearl - https://ftp.cs.ucla.edu/pub/stat_ser/r414.pdf

A covid example of collider bias: https://www.nature.com/articles/s41467-020-19478-2

"Fixed" and "Mixed" Models in DAG form



Glossary

Endogenous variable - a variable that is changed or determined by other variables (AKA a dependent variable)

Exogenous variable - independent variable

Collider - a third variable that is caused by two other variables along some path

Backdoor path - path between X and Y that contains an arrow into X
-can close a backdoor path by conditioning on any confounder/fork along the path
-backdoor path is closed if it contains a collider; conditioning on the collider opens the path

Average treatment effect (ATE) - Expected difference between potential outcome when treatment is applied and potential outcome when treatment is not applied

Independence assumption - Treatment assignment is independent of potential outcomes

Stable Unit Treatment Value Assumption (SUTVA) -

1) The potential outcomes for any unit do not vary with the treatments assigned to other units (no interference)
2) For each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes (no hidden variations of treatments)

Selection bias - expected difference between potential outcome in absence of treatment for the treatment group and potential outcome in absence of treatment for the untreated group (e.g., when people in the treatment group would have had better outcomes than the untreated group even if they hadn't received the treatment)

Heterogeneous treatment effect bias - arises when the expected treatment effect differs between the treated and untreated groups (e.g., when people who are more likely to benefit from a treatment are more likely to be in the treated group)

Backdoor criterion - rules for selecting a set of variables Z to condition on when trying to estimate the causal relationship between X and Y
Z satisfies the backdoor criterion if Z contains no descendents of X and Z blocks every path between X and Y that contains an arrow into X (i.e., backdoor)
To satisfy backdoor criterion, we want to

1) Block all spurious paths (backdoors) between X and Y
2) Leave all directed paths from X to Y unperturbed
3) Create no new spurious paths (e.g., by conditioning on a collider)