EMORY | COLLEGE OF ARTS AND SCIENCES

Causal Diagrams

Intro. to causal inference | SPSP 2023 Annual Convention

Wen Wei Loh (wen.wei.loh@emory.edu)
Department of Quantitative Theory and Methods (QTM)

February 25, 2023

Building blocks of causal diagrams

How can two variables be associated?

$$X \longrightarrow Y$$

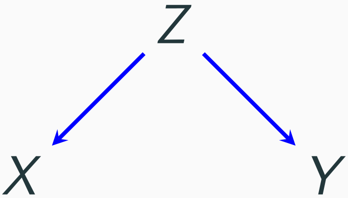
Two variables X and Y will be associated (or correlated) in the population if X causes Y; ...

How can two variables be associated?

$$X \longleftarrow Y$$

or if Y causes X; ...

How can two variables be associated?



or if Z is a common cause of both X and Y.

How can two variables be associated?

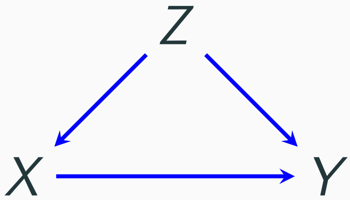
(i) (ii) (iii)

When X and Y are associated in the population, then at least one of the above must be true.

If X is a treatment or focal predictor of interest:

- which correspond(s) to a randomized experiment?
- which correspond(s) to an observational study?

How to remove an association due to Z?

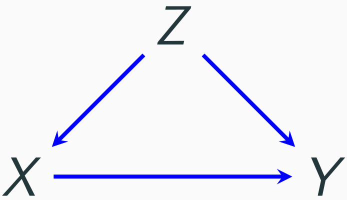


Our interest: causal effect of X on Y.

But Z generates a spurious (non-causal) $X - Y$ association.

I.e., Z induces confounding bias in the $X - Y$ estimate.

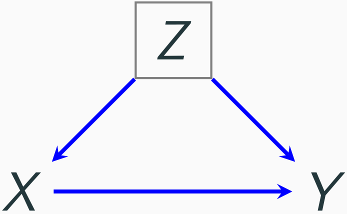
How to remove an association due to Z?



Adjusting for Z removes the association due to Z.

- E.g., *statistically control* for Z when regressing Y on X; or
- *condition* on Z by analyzing data with a specific value of Z;
- analyze X and Y by *stratifying* on Z;
- ...

How to remove an association due to Z?

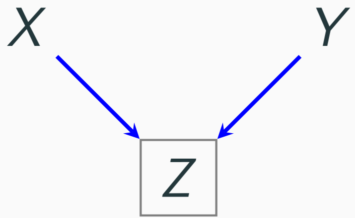


(Box around Z denotes adjusting for it.)

Observations are conditionally exchangeable given Z.

I.e., observations with the same value of Z are pseudo-randomized to different levels of X.

How can variables be associated in a sub-population?



Suppose X and Y are (unconditionally) independent.

Let Z be a common effect of both X and Y.

Then X and Y will be conditionally associated given Z.

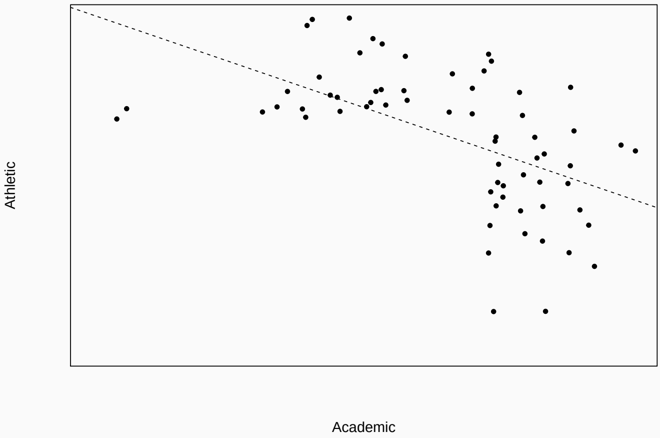
Adjusting for Z induces collider bias [Elwert and Winship, 2014, Greenland, 2003].

An example of collider bias

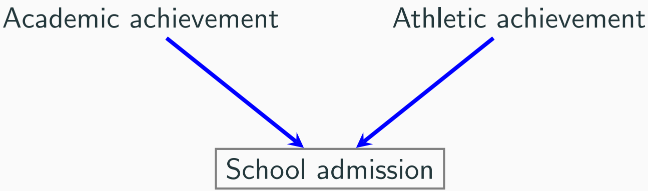
Suppose that within a certain school, academic achievement is negatively associated with athletic achievement.

Is there also a negative association in the general population?

An example of collider bias



An example of collider bias

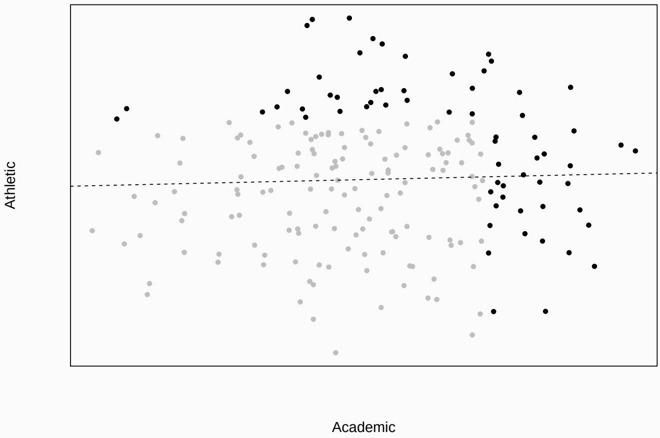


Not necessarily.

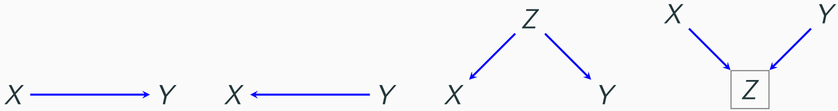
The school could have admitted students who excel in either academics or athletics.

If academic achievement and athletic achievement are independent in the population, then there would still be a negative association within the school.

An example of collider bias



Summary so far



(i) (ii) (iii) (iv)

X and Y will be associated under (i) – (iii).

X and Y will be associated after adjusting for Z under (iv).

Adjusting for Z removes confounding bias in (iii), but generates collider bias in (iv).

These are the building blocks of **causal diagrams**.

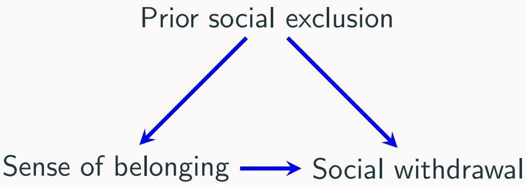
Examples

Draw a causal diagram for each scenario below, and determine whether there is any confounding or collider bias.

1. Interest is in the causal effect of reported sense of belonging on reported social withdrawal among first-year college students. Prior social exclusion is correlated with both sense of belonging and social withdrawal.
2. Interest is in the causal effect of hip fractures on myocardial infarction. Researchers collected data among hospitalized patients who could have been admitted for either hip fractures or myocardial infarction.

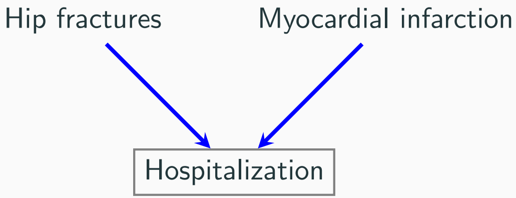
Examples

1. Prior social exclusion is a common cause of sense of belonging and social withdrawal.



Examples

2. This is an example of a *case-control* study: the cases were hospitalized patients with myocardial infarction, and the controls were hospitalized patients with hip fractures. Because the study was restricted to patients hospitalized for either reason, controls are more likely than cases to have had a hip fracture [Hernán et al., 2004].



Constructing causal diagrams

What are causal diagrams?

Causal diagrams[1] are graphical representations of the **data-generating mechanism or process**.

Visually represent knowledge as causal assumptions; e.g.,

- direction of causality;
- presence or absence of common causes;
- absence of causal effects;
- study design (e.g., sampling mechanisms).

They are nonparametric: do not depend on any statistical models.

[1] a.k.a. *causal graphs*, *causal Directed Acyclic Graphs (DAGs)*, or *causal Bayesian networks*

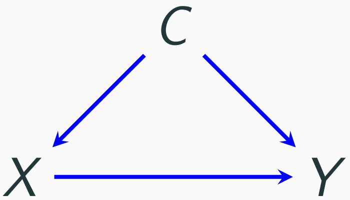
How to construct a causal diagram?

$$X \longrightarrow Y$$

Step 1

- State the treatment (X) and outcome (Y).
- Draw an arrow from X to Y.
- This arrow represents the causal effect we are targeting.
- Check: does this effect make sense conceptually?

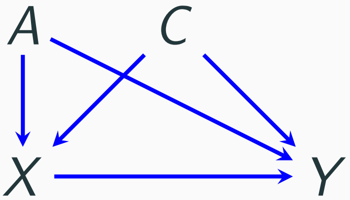
How to construct a causal diagram?



Step 2

- Add any known common cause C of X and Y.
- Draw arrows from C to X and C to Y.
- Include C irrespective of whether or not it has been measured.

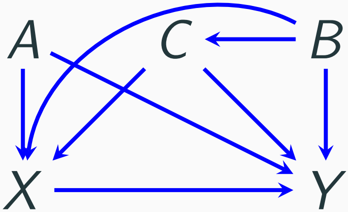
How to construct a causal diagram?



Step 2

- Continue adding any known common cause of X and Y.

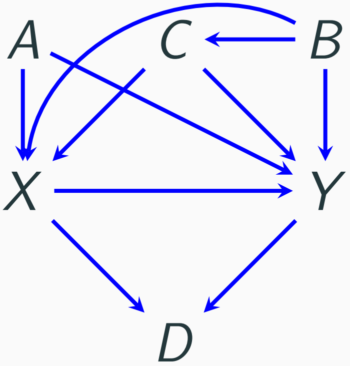
How to construct a causal diagram?



Step 3

- Add any known common cause of two or more variables already in the diagram.

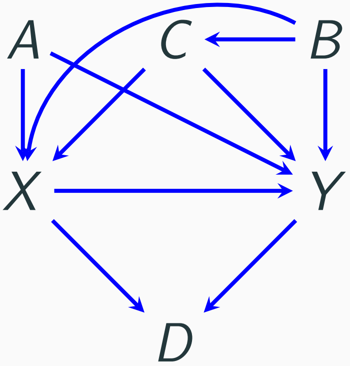
How to construct a causal diagram?



Step 4

- We can also add variables that are not common causes of others already in the diagram.
- E.g., D.

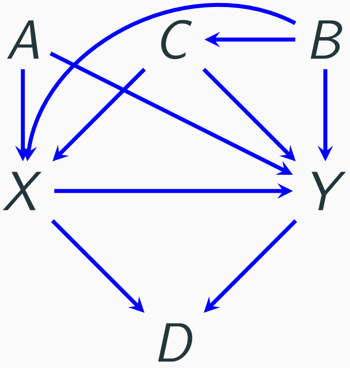
How to construct a causal diagram?



Remark

- The variables and arrows **not** in our diagram are part of our causal assumptions.
- E.g., A has no direct effect on B, C or D.
- E.g., B has no direct effect on A or D.

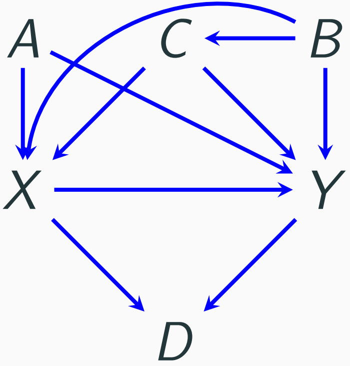
How to construct a causal diagram?



Remark

- A node or vertex on the diagram may represent a collection of variables; e.g., $A = (\text{age}, \text{gender})$.
- Each of these variables must satisfy the same causal relations!
- This avoids having to specify the relations among the collection.

Some terminology

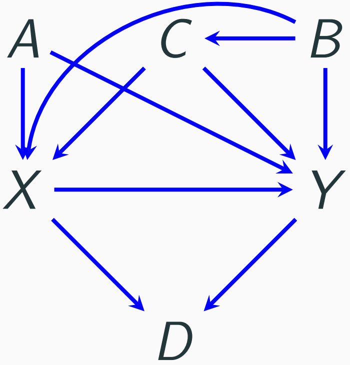


Path

- A path connecting two variables is a sequence of distinct variables connected by arrows.
- Paths from X to Y are, e.g.,

$$X \leftarrow C \rightarrow Y;$$
$$X \leftarrow C \leftarrow B \rightarrow Y;$$
$$X \rightarrow D \leftarrow Y.$$

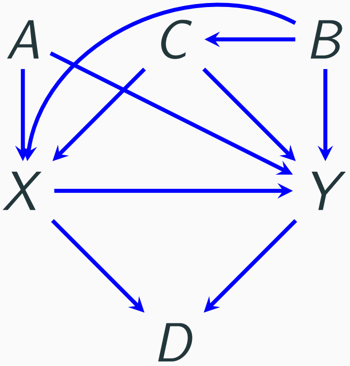
Some terminology



Descendants

- A directed path is a path with arrows all pointing in the same direction.
- E.g., which are the directed paths from X to D?
- A vertex, e.g., D, is a causal descendant of another vertex, e.g., X, if there is a directed path from X to D.

Some terminology

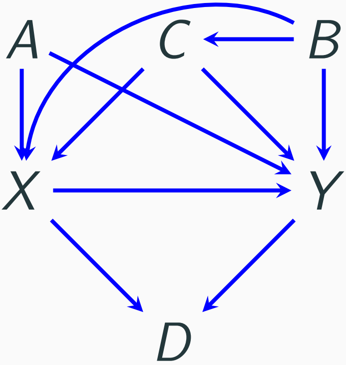


Collider

- A collider on a specific path has two arrows pointing directly at it.
- E.g., D is a collider on the path $X \rightarrow D \leftarrow Y$.
- **A collider is path-specific.**
- A *non-collider* is not a collider.

Detecting confounding and collider bias in causal diagrams

Detecting confounding and collider bias in causal diagrams



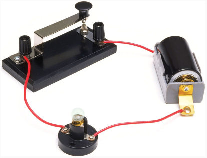
What next?

- Given a causal diagram, now determine if the $X \to Y$ relationship is prone to confounding or collider biases.
- This requires checking all possible paths with X and Y as the endpoints.

Checking open and closed paths

Think of a causal diagram as an electrical circuit.

- Open paths conduct (or transmit) electric current along them.
- Open paths generate **spurious** associations between the endpoints.



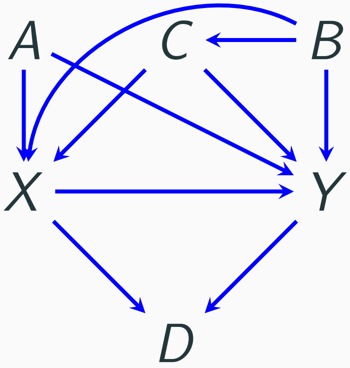
From: https://www.britannica.com/technology/electric-circuit.

Checking open and closed paths

- Closed paths stop the electric current.
- Closed paths stem spurious associations between the endpoints.

Goal: Check whether all paths linking X and Y, except the $X \rightarrow Y$ path, can be closed.

Checking open and closed paths



Rules (without adjustment)

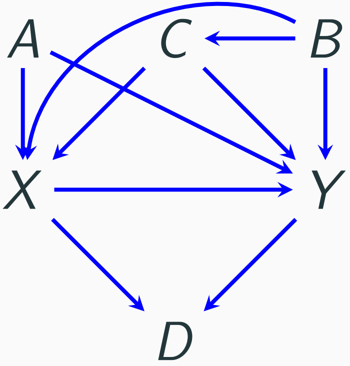
- Paths with non-colliders only are open.

$$X \leftarrow A \rightarrow Y;$$
$$X \leftarrow C \rightarrow Y;$$
$$X \leftarrow B \rightarrow Y;$$
$$X \leftarrow C \leftarrow B \rightarrow Y.$$

- Paths with a collider are closed.

$$X \rightarrow D \leftarrow Y.$$

Checking open and closed paths



Rules (with adjustment)

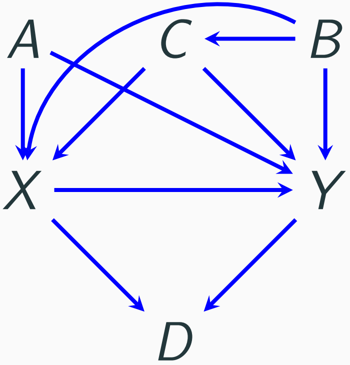
- Adjusting for a non-collider along a path closes the path.

$$X \leftarrow \boxed{A} \rightarrow Y;$$
$$X \leftarrow \boxed{C} \rightarrow Y;$$
$$X \leftarrow \boxed{B} \rightarrow Y;$$
$$X \leftarrow \boxed{C} \leftarrow \boxed{B} \rightarrow Y.$$

- Adjusting for a collider or a descendant of a collider opens the path.

$$X \rightarrow \boxed{D} \leftarrow Y.$$

Checking open and closed paths



Which variables to adjust for?

- Can we adjust for a subset of variables to close all paths linking X and Y (except the $X \to Y$ path)?

$$X \leftarrow \boxed{A} \rightarrow Y;$$

$$X \leftarrow \boxed{C} \rightarrow Y;$$

$$X \leftarrow \boxed{B} \rightarrow Y;$$

$$X \leftarrow \boxed{C} \leftarrow \boxed{B} \rightarrow Y;$$

$$X \rightarrow D \leftarrow Y.$$

Exercise 1

$$
\begin{tikzcd}
A \arrow[r] \arrow[d] \arrow[drr] & C & B \arrow[l] \arrow[d] \\
X \arrow[rr] & & Y
\end{tikzcd}
$$

Which variables to adjust for?

Exercise 1

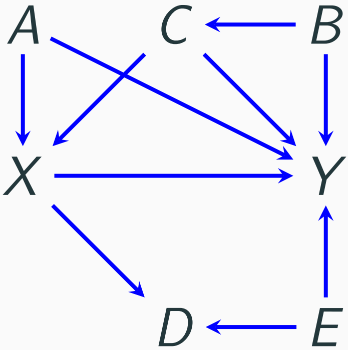
$$
\begin{tikzcd}
A \arrow[r] \arrow[d] \arrow[drr] & C & B \arrow[l] \arrow[d] \\
X \arrow[rr] & & Y
\end{tikzcd}
$$

Which variables to adjust for?

$$X \leftarrow \boxed{A} \rightarrow Y;$$
$$X \leftarrow \boxed{A} \rightarrow C \leftarrow B \rightarrow Y.$$

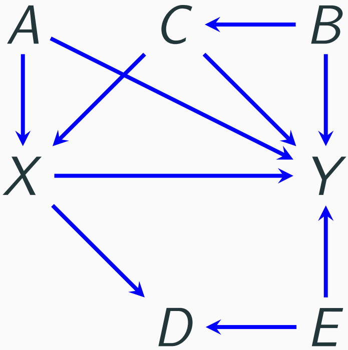
- What about C?
- What about B?

Exercise 2



Which variables to adjust for?

Exercise 2



Which variables to adjust for?

$X \leftarrow \boxed{A} \rightarrow Y;$

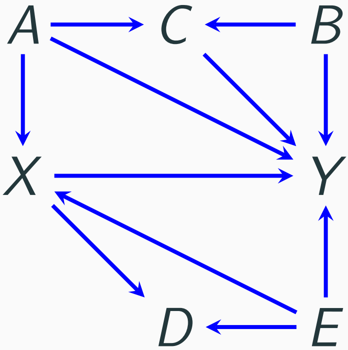
$X \leftarrow \boxed{C} \rightarrow Y;$

$X \leftarrow \boxed{C} \leftarrow B \rightarrow Y;$

$X \rightarrow D \leftarrow E \leftarrow Y.$

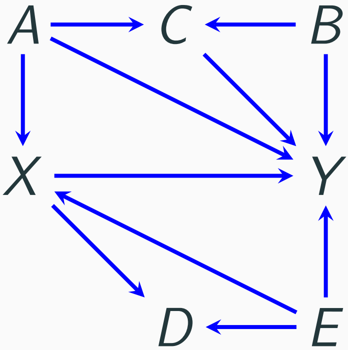
- What about B?
- What about D?
- What about E?

Exercise 3



Which variables to adjust for?

Exercise 3



Which variables to adjust for?

$X \leftarrow \boxed{A} \rightarrow Y;$

$X \leftarrow \boxed{A} \rightarrow C \rightarrow Y;$

$X \leftarrow \boxed{A} \rightarrow C \leftarrow B \rightarrow Y;$

$X \leftarrow \boxed{E} \rightarrow Y;$

$X \rightarrow D \leftarrow \boxed{E} \leftarrow Y.$

- What about B?
- What about C?
- What about D?

Formally: Back-door criterion

- *Unconfoundedness* assumption: all confounding (and collider) biases can be eliminated from a causal effect by adjusting for a covariate subset.
- *Back-door criterion*: a set of graphical conditions for determining whether the unconfoundedness assumption can be met [Hayduk et al., 2003, Pearl, 2009].
  1. All paths linking treatment and outcome, with an arrow pointing to treatment (i.e., "back-door" paths), must be closed after adjusting for the subset.
  2. No variable in the subset is a causal descendant of treatment.

Formally: Back-door criterion

- Challenging to enumerate all possible paths, then check each candidate subset in practice.
- DAGitty: Automated tool to visualize causal diagrams [Textor et al., 2017].
- Determine whether the back-door criterion can be met by adjusting for a (minimal) covariate subset.
  - Exercise 2: `http://dagitty.net/mevvRRK`
  - Exercise 3: `http://dagitty.net/mx9ZuN7`

Conclusion

Conclusion

- Causal inference requires careful consideration of confounding and collider biases.
- Causal diagrams are intuitive visual representations of substantive knowledge [Elwert, 2013, Glymour et al., 2016, Grosz et al., 2020, Morgan and Winship, 2015, Rohrer, 2018].
- They encode causal assumptions about "what may (or does not) exert a causal effect on what."
- When estimating a causal effect, can use graphical rules to determine:
  - which variables to adjust for to eliminate confounding bias; and
  - which variables not to adjust for to avoid collider bias.

References

F. Elwert. *Graphical Causal Models*, pages 245–273. Springer Netherlands, Dordrecht, 2013. ISBN 978-94-007-6094-3. doi: 10.1007/978-94-007-6094-3_13. URL https://doi.org/10.1007/978-94-007-6094-3_13.

F. Elwert and C. Winship. Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40(1):31–53, 2014. doi: 10.1146/annurev-soc-071913-043455. URL https://doi.org/10.1146/annurev-soc-071913-043455. PMID: 30111904.

M. M. Glymour, N. P. Jewell, and J. Pearl. *Causal inference in statistics: a primer*. John Wiley & Sons Ltd, Chichester, West Sussex, UK, 2016. ISBN 9781119186847.

S. Greenland. Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology*, 14(3):300–306, 2021/08/13/ 2003. URL http://www.jstor.org/stable/3703850.

M. P. Grosz, J. M. Rohrer, and F. Thoemmes. The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, 15(5):1243–1255, 2020. doi: 10.1177/1745691620921521. URL https://doi.org/10.1177/1745691620921521. PMID: 32727292.

L. Hayduk, G. Cummings, R. Stratkotter, M. Nimmo, K. Grygoryev, D. Dosman, M. Gillespie, H. Pazderka-Robinson, and K. Boadu. Pearl's D-separation: One more step into causal thinking. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(2):289–311, 2003. doi: 10.1207/s15328007sem1002_8.

M. A. Hernán, S. Hernández-Díaz, and J. M. Robins. A structural approach to selection bias. *Epidemiology*, 15(5), 2004. URL https://journals.lww.com/epidem/Fulltext/2004/09000/A_Structural_Approach_to_Selection_Bias.20.aspx.

S. L. Morgan and C. Winship. *Counterfactuals and causal inference*. Cambridge University Press, 2015.

J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009. doi: 10.1017/CBO9780511803161.

J. M. Rohrer. Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1):27–42, 2018. doi: 10.1177/2515245917745629. URL https://doi.org/10.1177/2515245917745629.

J. Textor, B. van der Zander, M. S. Gilthorpe, M. Liśkiewicz, and G. T. Ellison. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology*, 45(6):1887–1894, 01 2017. ISSN 0300-5771. doi: 10.1093/ije/dyw341. URL https://doi.org/10.1093/ije/dyw341.