Introduction to Directed Acyclic Graphs (DAGs) for Causal Inference

Session 1

Ione Avila-Palencia, PhD, MPH

Hüseyin Küçükali, MD, PhD

11th May 2023

This work was supported by **UK Research and Innovation** [ES/V016075/1]

Economic and Social Research Council

Healthy Ageing Challenge Social, Behavioural and Design Research

Facilitators

Dr Ione Avila-Palencia
Research Fellow
Queen's University Belfast
i.avila-palencia@qub.ac.uk
@ioneavpa

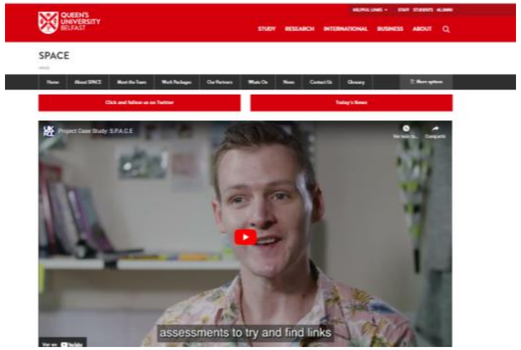
Dr Hüseyin Küçükali
Research Fellow
Queen's University Belfast
h.kucukali@qub.ac.uk
@hkucukali

SPACE Project

SUPPORTIVE ENVIRONMENTS FOR PHYSICAL & SOCIAL ACTIVITY, HEALTHY AGEING & COGNITIVE HEALTH

- Aim:
  - **to investigate the impacts, and possible mechanistic pathways, of urban environments on healthy ageing and the cognitive health**
- This research builds on several projects:
- Interdisciplinary project
- More info:

qub.ac.uk/sites/space/
@spacequb

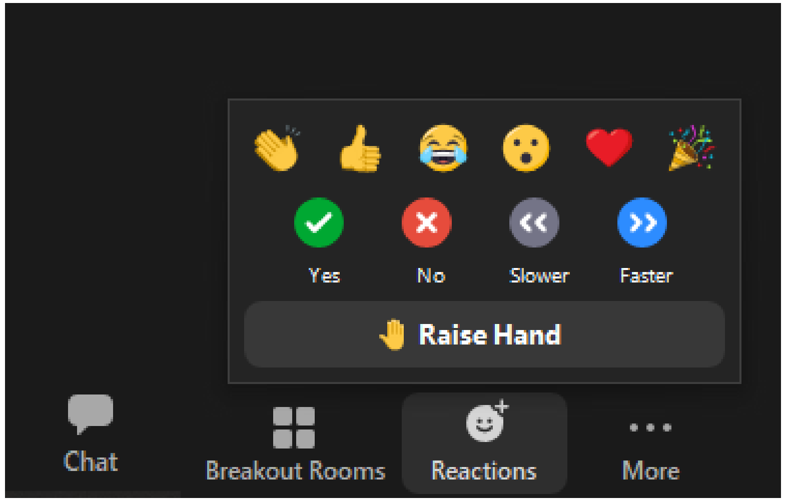


Outline

- Short Introduction to causality
- DAGs: the essentials
- Confounder, mediator, and collider
- SPACE example
- d-separation rules
- Work with examples
- DAGitty demonstration

Webinar participation

- Participants can use the Q&A to ask questions
- Questions breaks - Open to general question throughout (Q&A, raise hand)



3 Roles of Data Science

Description

- what
- who

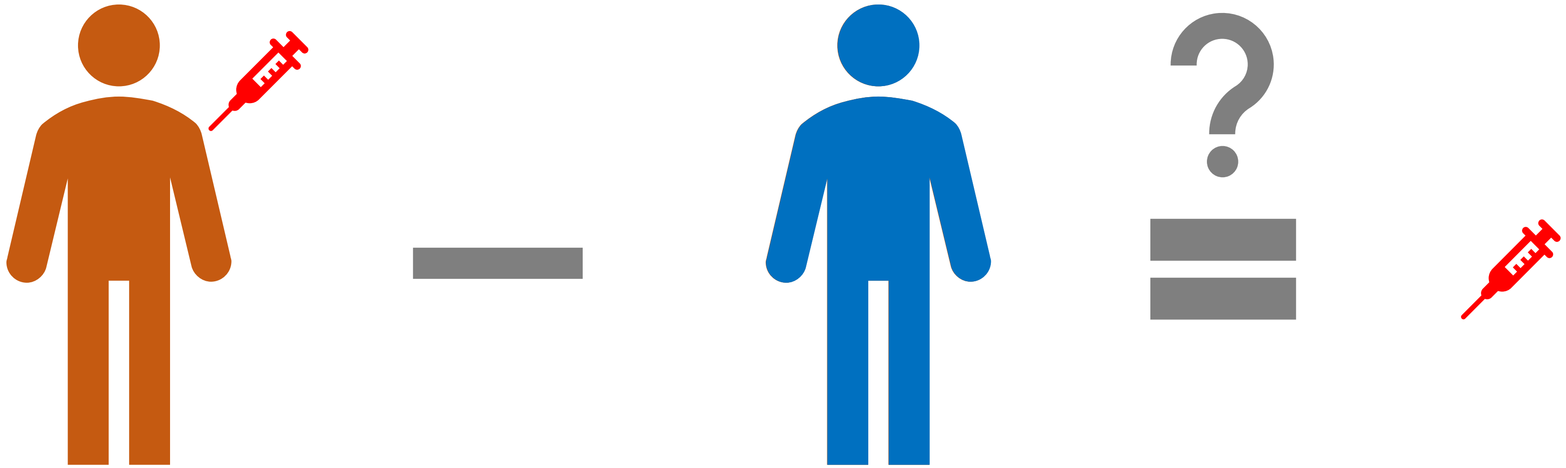
Prediction

- what will
- who will

Causal Inference

- why
- what-if

Counterfactual Thinking



Fundamental problem of causal inference

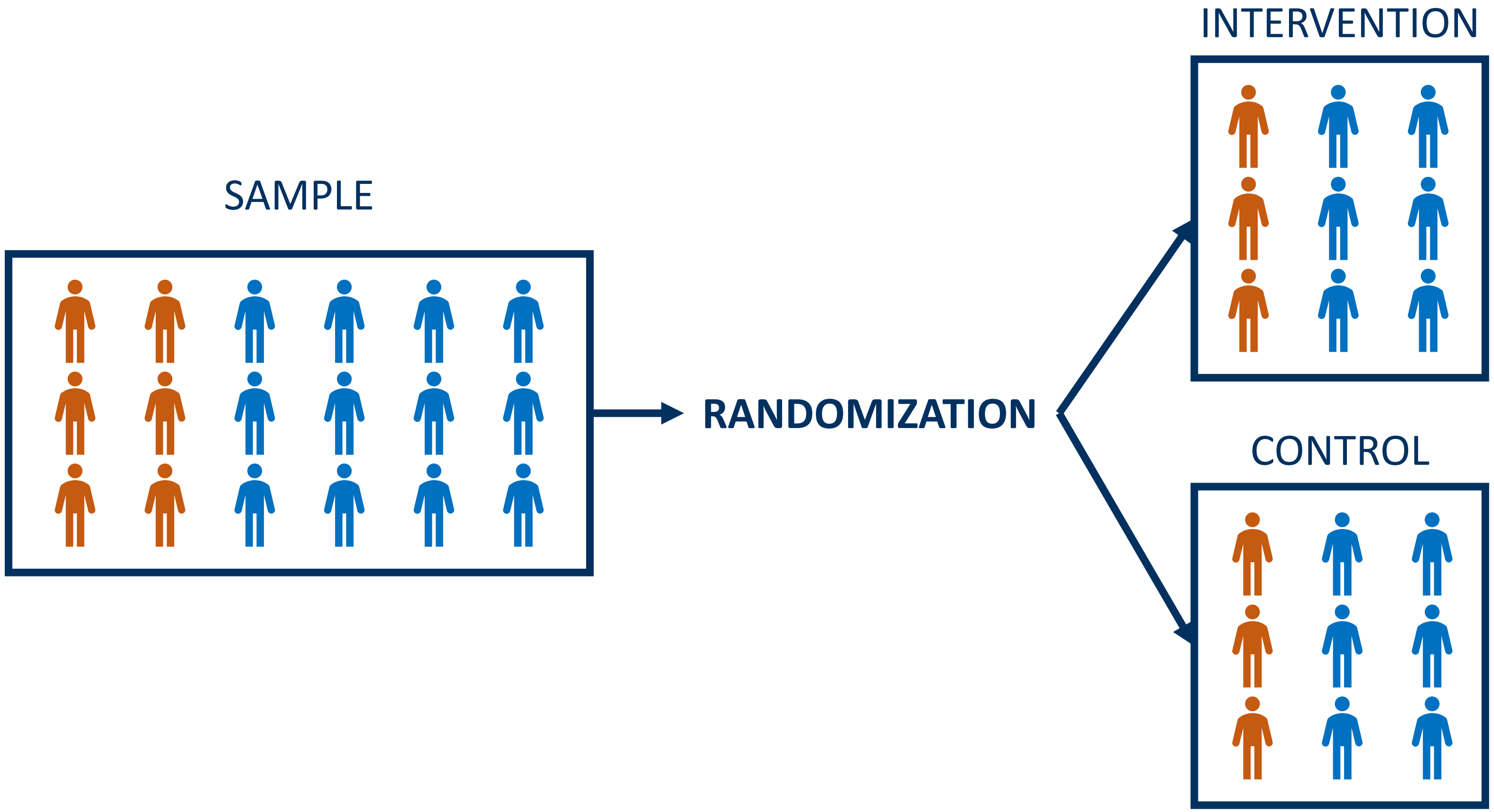
Exchangeability

Although we can not achieve exchangablity at individual level we can do so at group level.

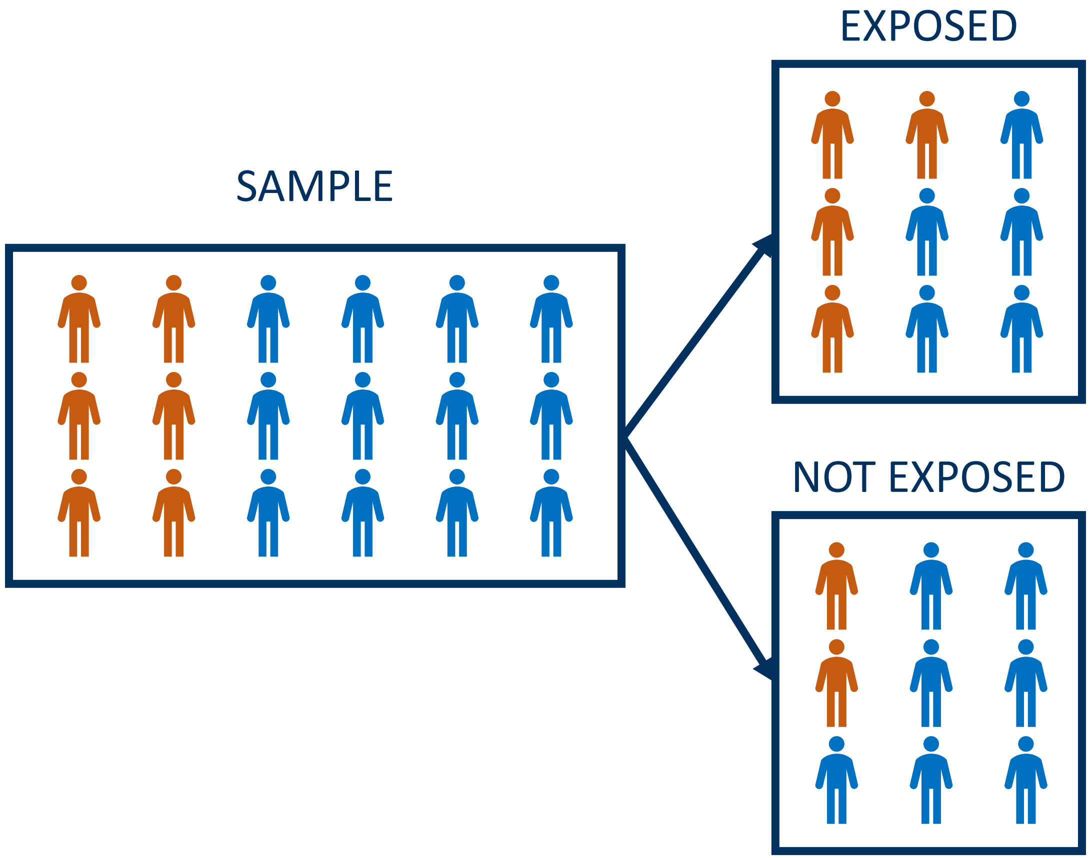
Unconditional Exchangeability

Conditional Exchangeability

Unconditional Exchangeability



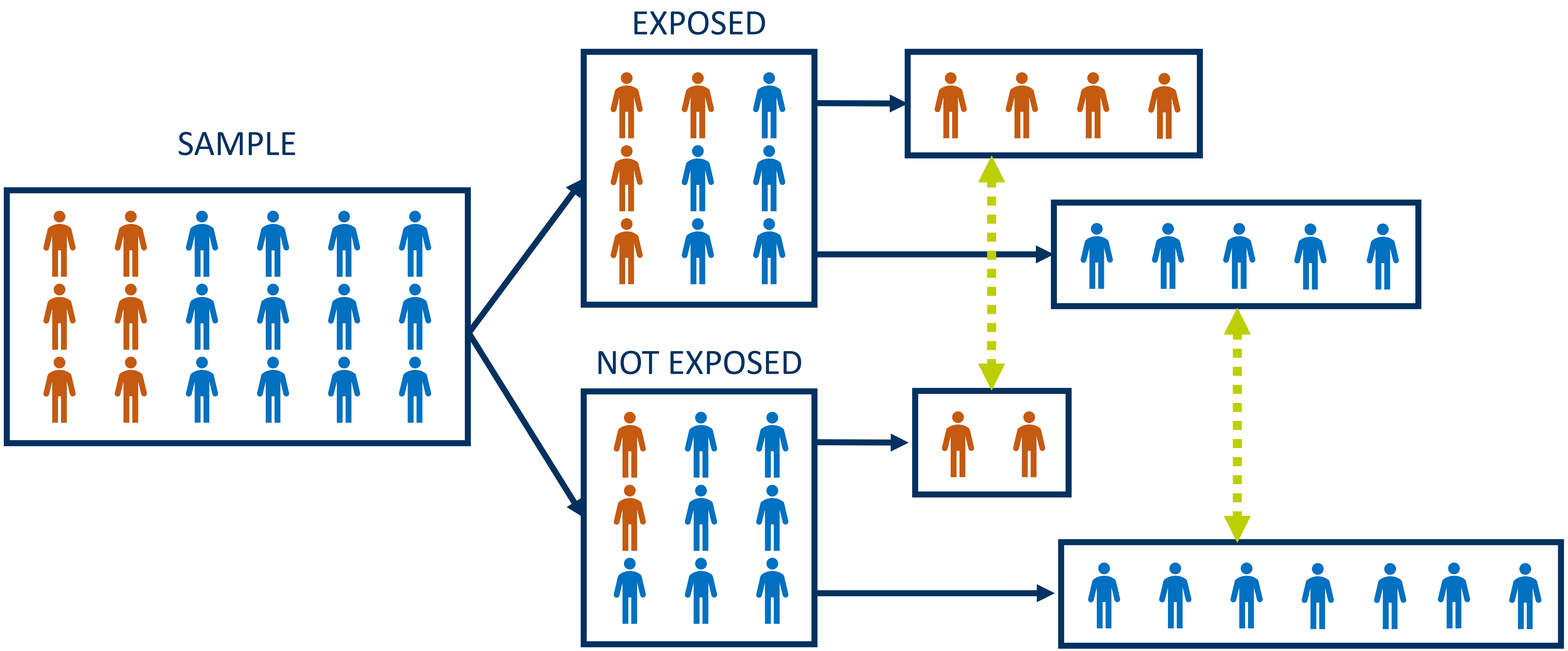
What about Observational Data?



Major problems

- Confounder Bias
- Collider Bias
- Selection Bias

Conditional Exchangeability



Why do we need DAGs?

We need to know the data generation process

- To understand
- To explain
- To avoid biases

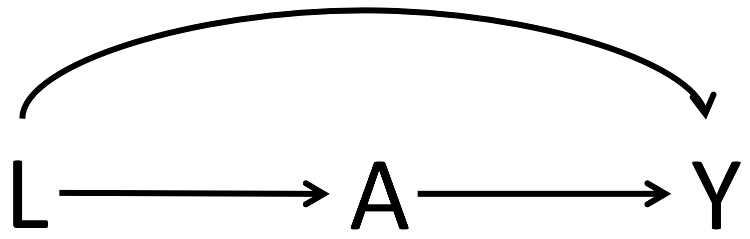
Questions

What is a DAG?

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

- DAG: Directed Acyclic Graph
- Mathematical object built from letters and arrows
- Visual representation of qualitative causal assumptions

What is a DAG useful for?



- Study design – e.g., dealing with selection bias
- Analysis plan – e.g., decisions on confounders
- Results interpretation
- Communication with other researchers
- Reading scientific literature

Elements of a DAG

- Nodes (vertices): variables that may be observed or unobserved

A Y

In this class we will use the following:
- A will represent exposure (or treatment)
- Y will represent outcome
- L will represent other measured variables
- U will represent other unmeasured variables

Elements of a DAG

- Edge

Elements of a DAG

- Arrow (directed edge): possible direct causal effect

Elements of a DAG

$$A \longrightarrow Y$$

A has a causal effect on Y
OR
We don’t know whether A has a causal effect on Y or not

Complete DAGs do not exclude any possible causal effect

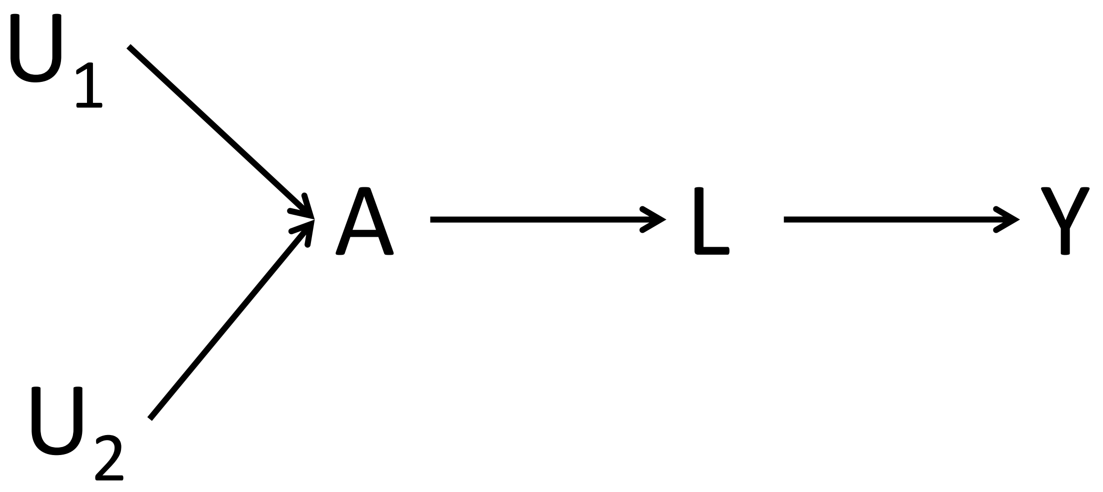
$$A \qquad\qquad\qquad Y$$

A does not have a causal effect on Y

Incomplete DAGs encode expert knowledge in the form of missing arrows

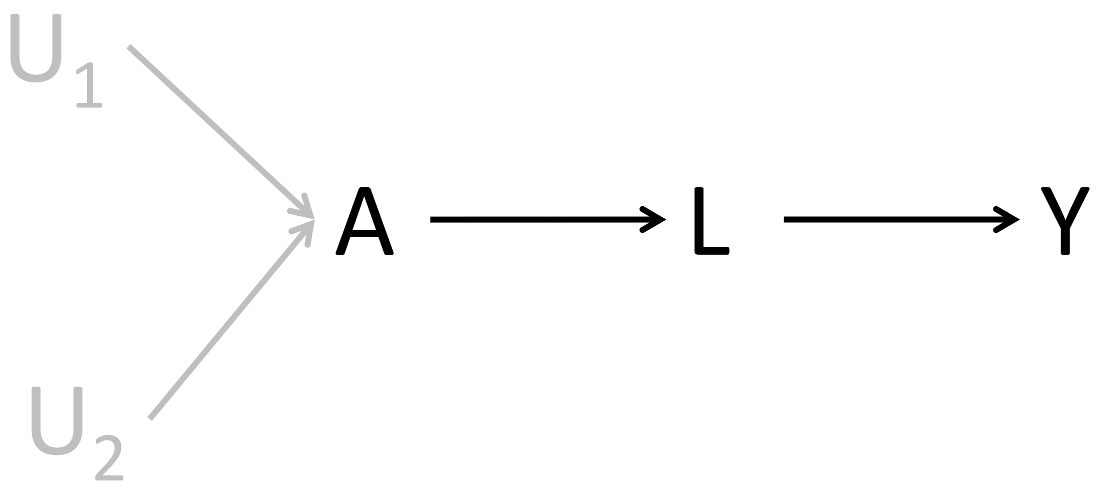
Elements of a DAG

- Node terminology



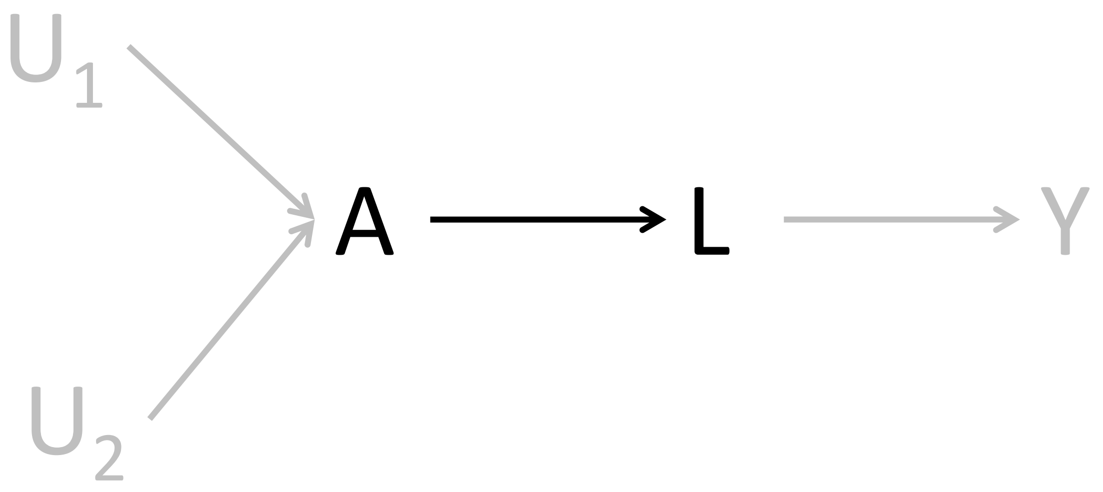
Elements of a DAG

- Node terminology
  - Descendants: all nodes directly or indirectly caused by the node. Desc (A)={L,Y}



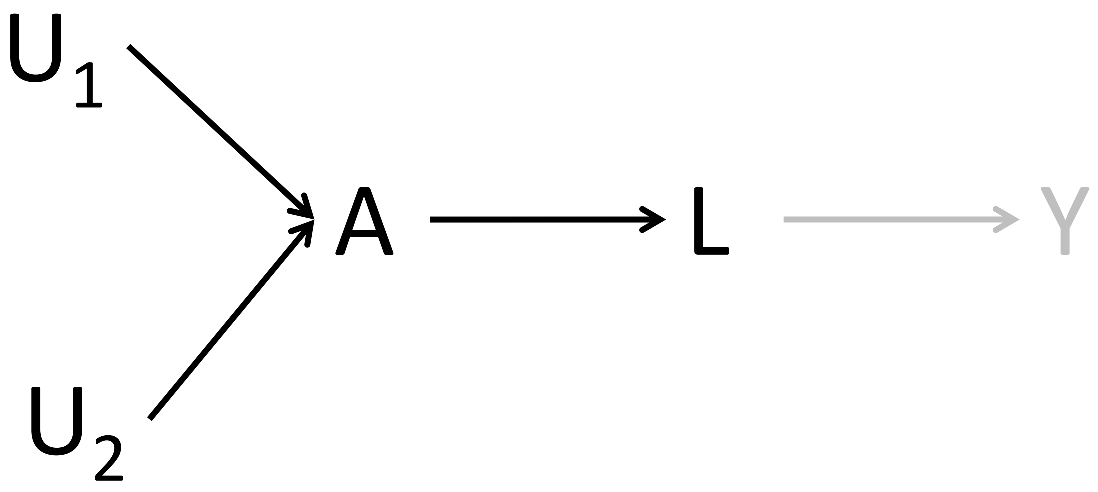
Elements of a DAG

- Node terminology
  - Descendants: all nodes directly or indirectly caused by the node. Desc (A)={L,Y}
  - Children: all nodes directly caused by the node. Child (A)={L}



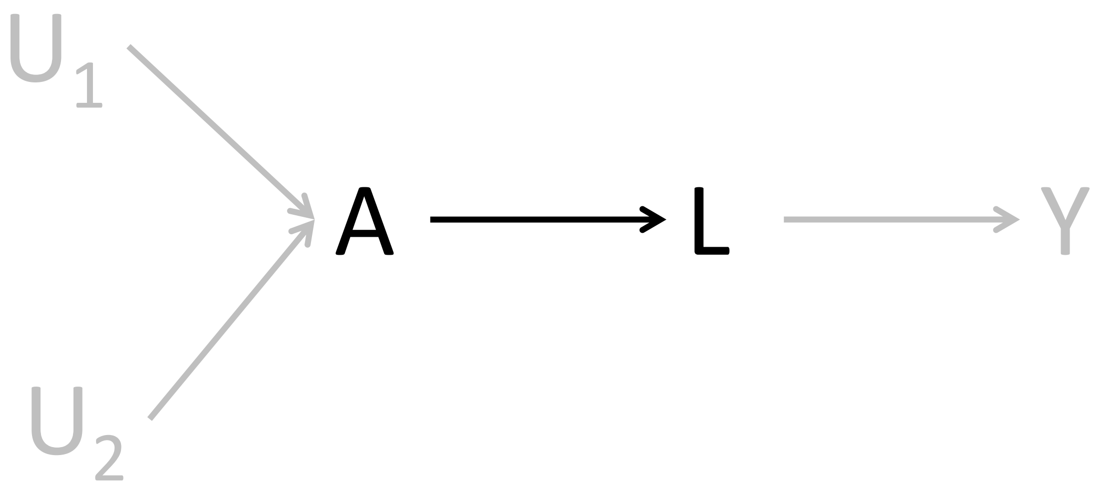
Elements of a DAG

- Node terminology
  - Descendants: all nodes directly or indirectly caused by the node. Desc (A)={L,Y}
  - Children: all nodes directly caused by the node. Child (A)={L}
  - Ancestors: all nodes directly or indirectly causing the node. An (L) = {A, U1, U2}



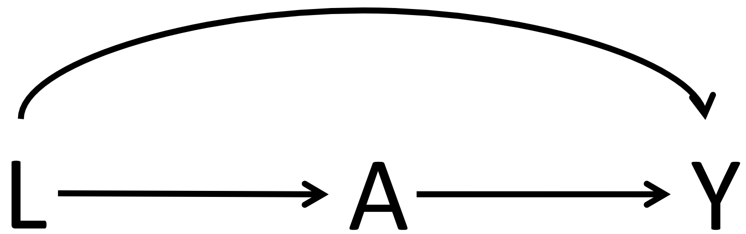
Elements of a DAG

- Node terminology
  - Descendants: all nodes directly or indirectly caused by the node. Desc (A)={L,Y}
  - Children: all nodes directly caused by the node. Child (A)={L}
  - Ancestors: all nodes directly or indirectly causing the node. An (L) = {A, U1, U2}
  - Parents: all direct causes of the node. Pa (L)= {A}

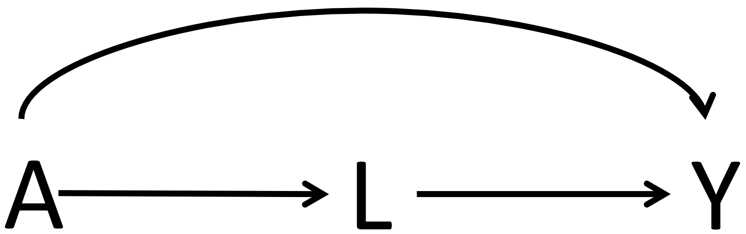


What is a DAG?

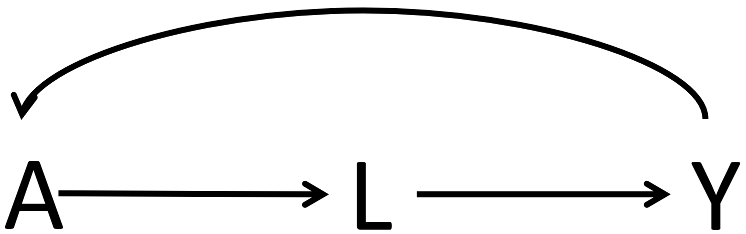
Are they Directed?
Are they Acyclic?
Are they Graphs?



This is a DAG

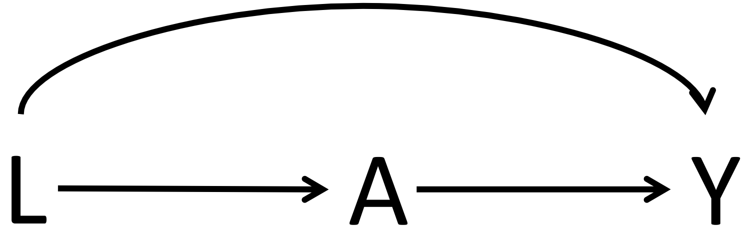


This is a DAG

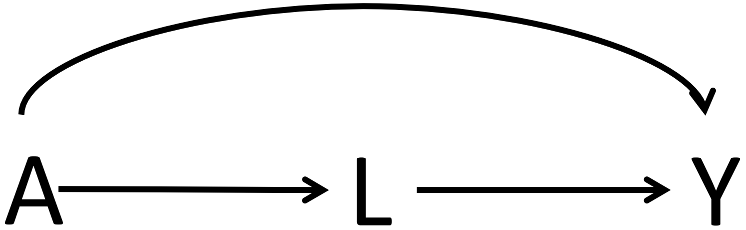


This is NOT a DAG

What is a DAG?

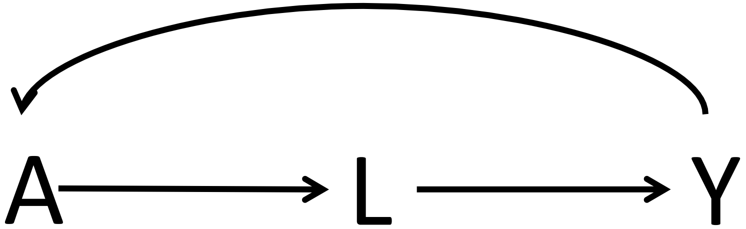


This is a DAG



This is a DAG

- DIRECTED edges (arrows) linking nodes (variables)
- ACYCLIC links because no arrows from descendants (effects) to ancestors (causes)
- GRAPHS



This is NOT a DAG

What is a DAG?

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

This is a DAG

$$\begin{tikzcd} A \arrow[r] \arrow[rr, bend left] & L \arrow[r] & Y \end{tikzcd}$$

This is a DAG

- DIRECTED edges (arrows) linking nodes (variables)
- ACYCLIC links because no arrows from descendants (effects) to ancestors (causes)
- GRAPHS

- 3 nodes, representing 3 random variables (L, A, Y)
- Arrows, representing causal effects
- By convention: time goes from left to right

Confounder, mediator, collider

Confounder, mediator, collider

Confounder

(L common cause A, Y)

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A & Y \end{tikzcd}$$

Confounder, mediator, collider

Confounder

(L common cause A, Y)

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A & Y \end{tikzcd}$$

Mediator

(L between A, Y)

$$\begin{tikzcd} A \arrow[r] & L \arrow[r] & Y \end{tikzcd}$$

Confounder, mediator, collider

Confounder

(L common cause A, Y)

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A & Y \end{tikzcd}$$

Mediator

(L between A, Y)

$$\begin{tikzcd} A \arrow[r] & L \arrow[r] & Y \end{tikzcd}$$

Collider

(L common effect A, Y)

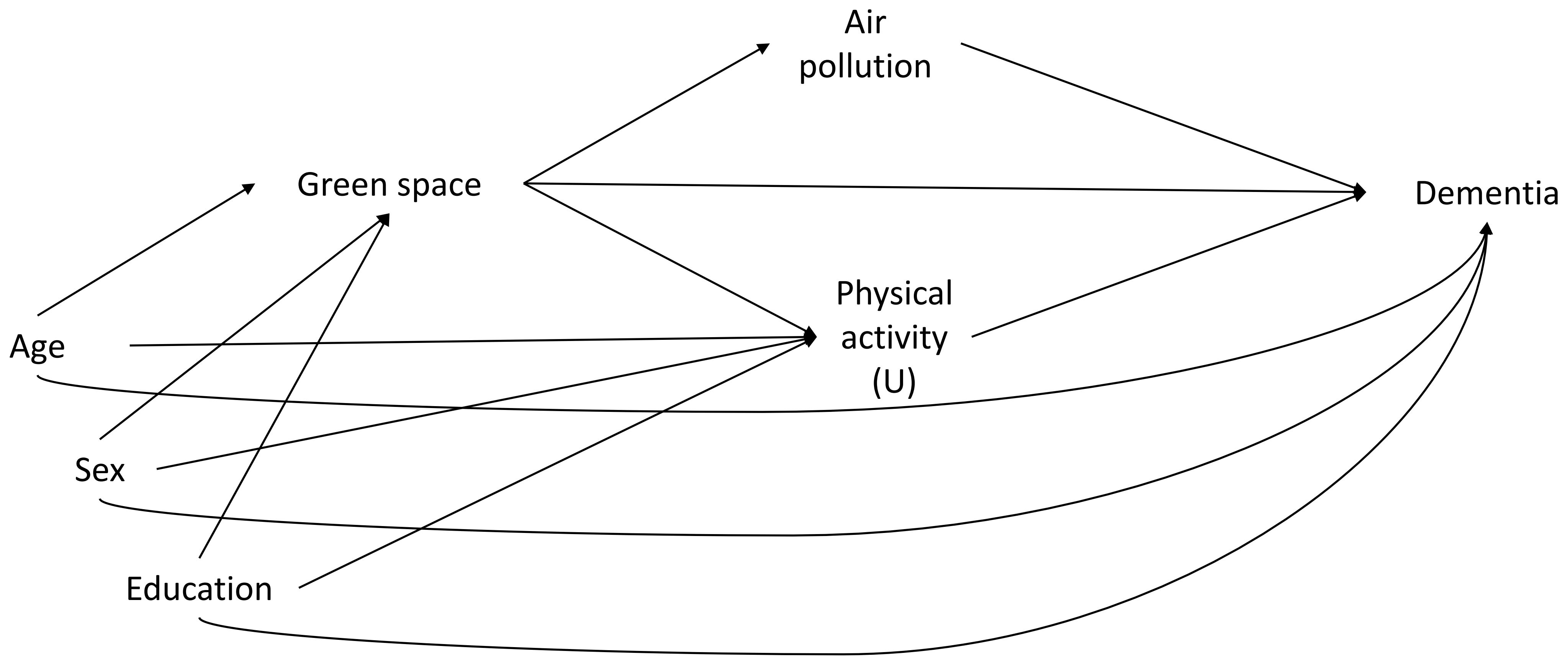
$$\begin{tikzcd} A \arrow[rr, bend left] & Y \arrow[r] & L \end{tikzcd}$$

Questions

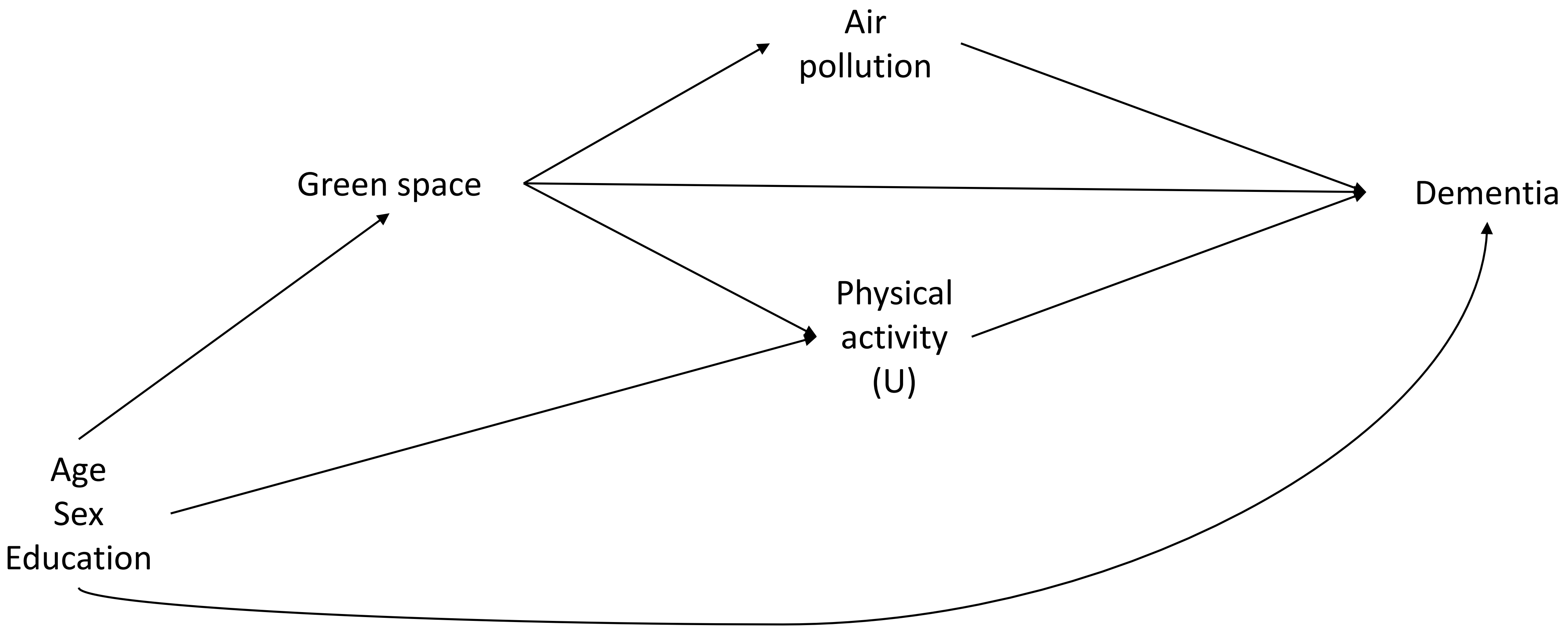
SPACE example

- Is green space associated with dementia?
  - Exposure: green space
  - Outcome: dementia
  - Covariates: age, sex, education, physical activity (unmeasured), air pollution

SPACE example



SPACE example



SPACE example

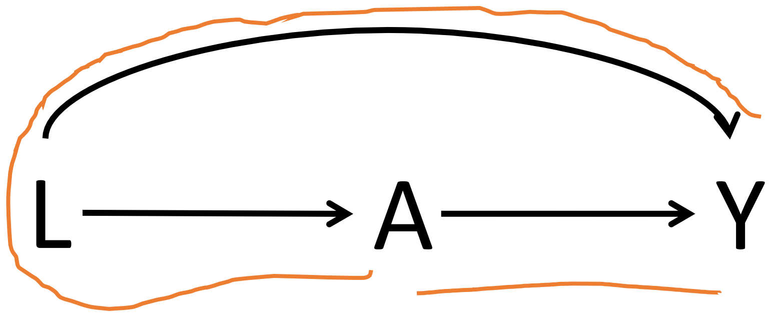
- Is green space associated with dementia?
  - Exposure: green space
  - Outcome: dementia
  - Covariates: age, sex, education, physical activity (unmeasured), air pollution
    - Confounders: age, sex, education
    - Mediators: physical activity (unmeasured), air pollution
    - Colliders (path specific): physical activity (unmeasured)

(green space and dementia could be considered colliders too, but as they are our exposure and outcome we won’t focus on them)

Questions

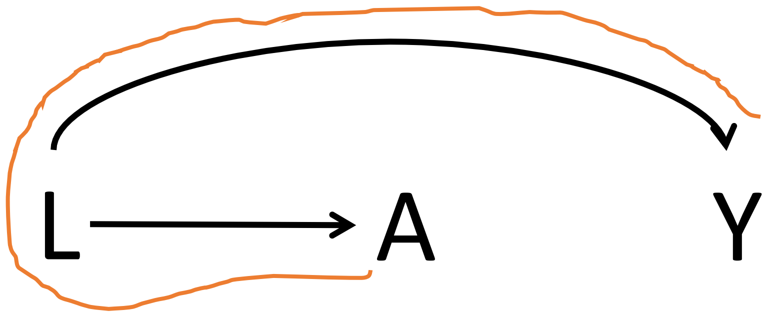
Paths

Paths from A to Y



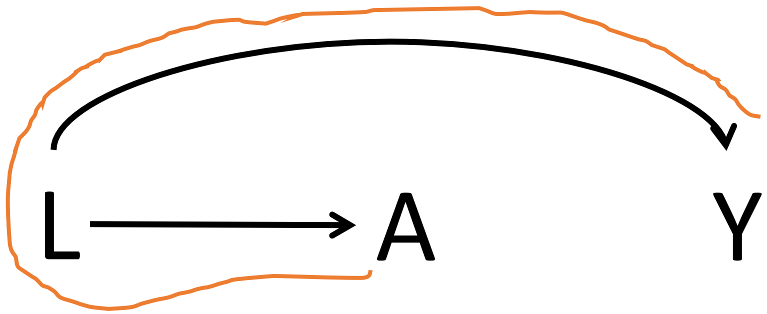
- A path is a sequence to get from one vertex (variable) to another traveling between edges (arrows)
- Rules:
  - Cannot pass by the same vertex (variable) twice
  - Can move against the direction of the arrow (backdoor path)

Paths

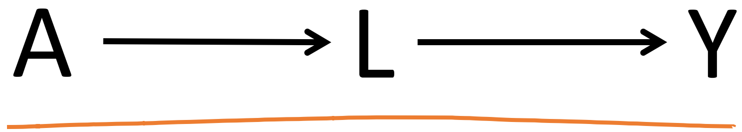


A ← L → Y   Non-causal path

Paths



A ← L → Y   Non-causal path



A → L → Y   Causal path

Paths

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A & Y \end{tikzcd}$$

A ← L → Y   Non-causal path

$$\begin{tikzcd} A \arrow[r] & L \arrow[r] & Y \end{tikzcd}$$

A → L → Y   Causal path

$$\begin{tikzcd} A \arrow[rr, bend left] & Y \arrow[r] & L \end{tikzcd}$$

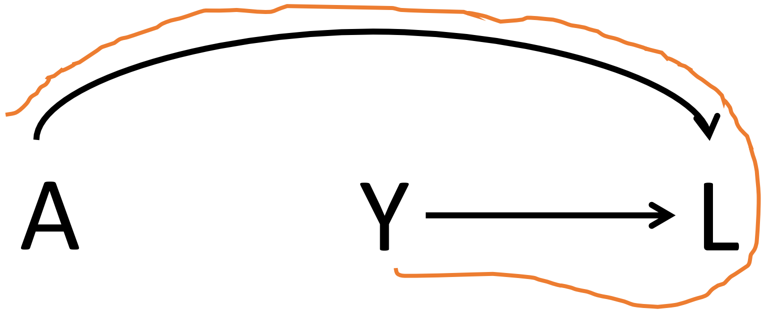
A → L ← Y   Non-causal path, collider

SPACE
QUEEN'S UNIVERSITY BELFAST

d-separation rules (Pearl, 1995)

d-separation rules (Pearl, 1995)

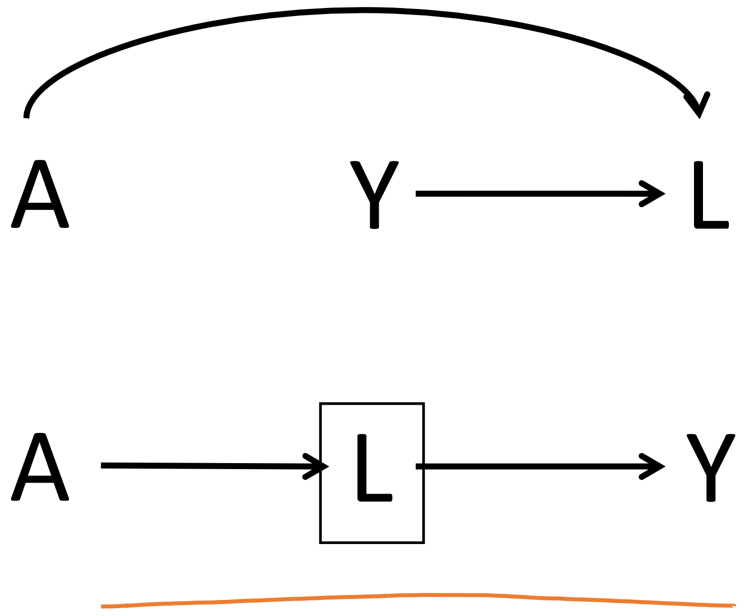
RULE #1: If there are no variables being conditioned on, <u>a path is blocked</u> if and only if two arrowheads on the path collide at some variable on the path.



d-separation rules (Pearl, 1995)

RULE #1: If there are no variables being conditioned on, a path is blocked if and only if two arrowheads on the path collide at some variable on the path.

RULE #2: Any path that contains a non-collider that has been conditioned on is blocked.

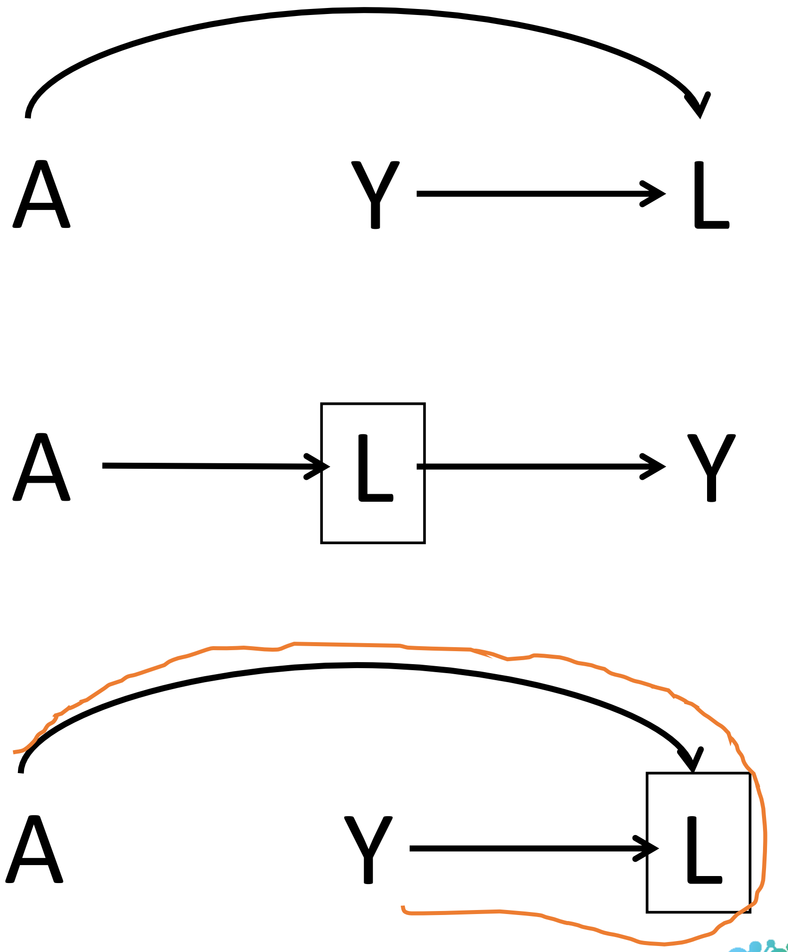


d-separation rules (Pearl, 1995)

RULE #1: If there are no variables being conditioned on, a path is blocked if and only if two arrowheads on the path collide at some variable on the path.

RULE #2: Any path that contains a non-collider that has been conditioned on is blocked.

RULE #3: A collider that has been conditioned on does not block a path.



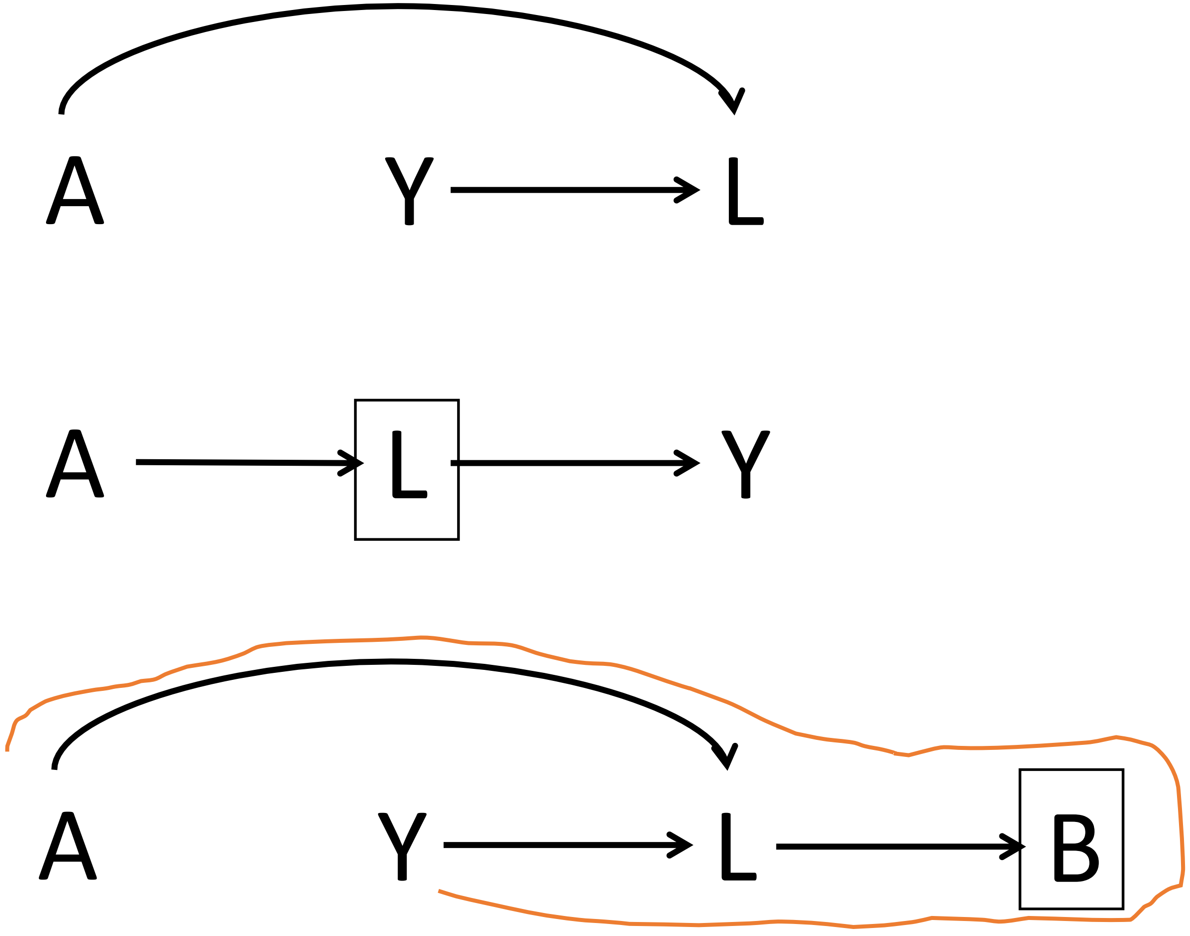
d-separation rules (Pearl, 1995)

RULE #1: If there are no variables being conditioned on, a path is blocked if and only if two arrowheads on the path collide at some variable on the path.

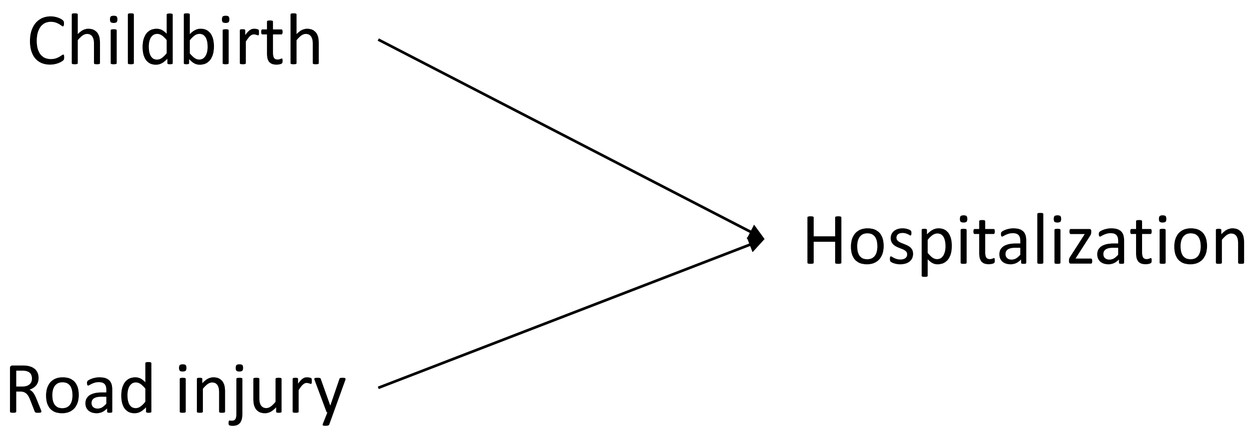
RULE #2: Any path that contains a non-collider that has been conditioned on is blocked.

RULE #3: A collider that has been conditioned on does not block a path.

RULE #4: A collider that has a descendant that has been conditioned on does not block a path.

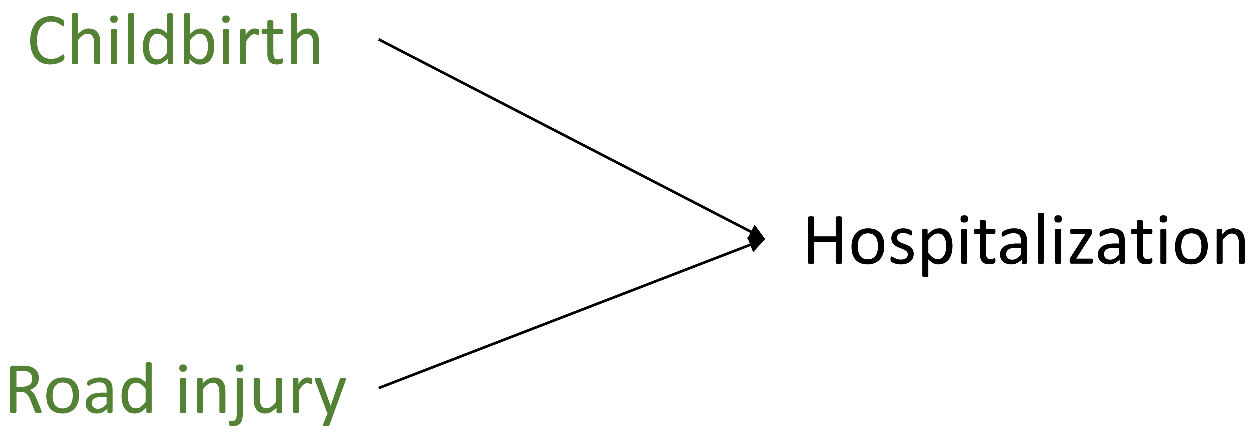


Collider example



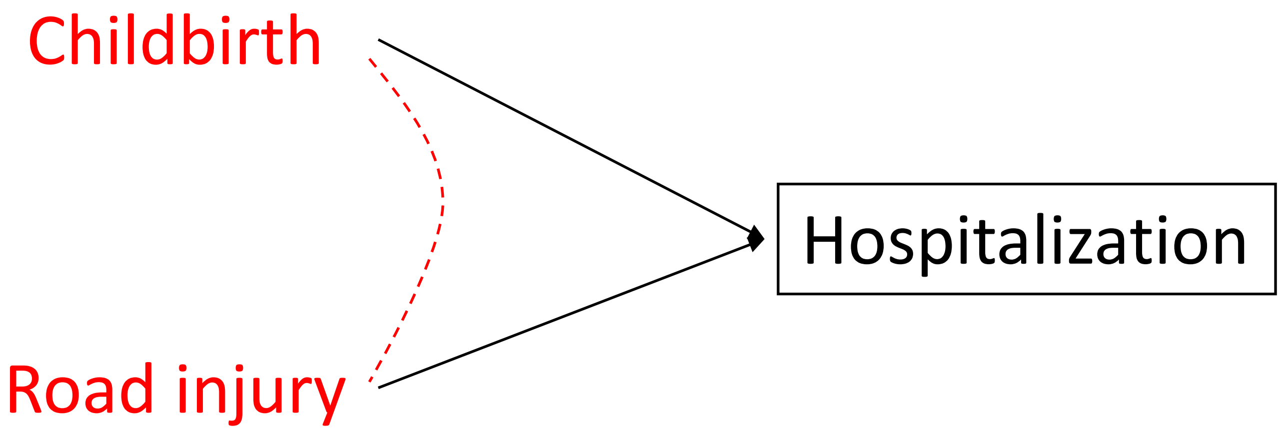
Collider example

d-separated

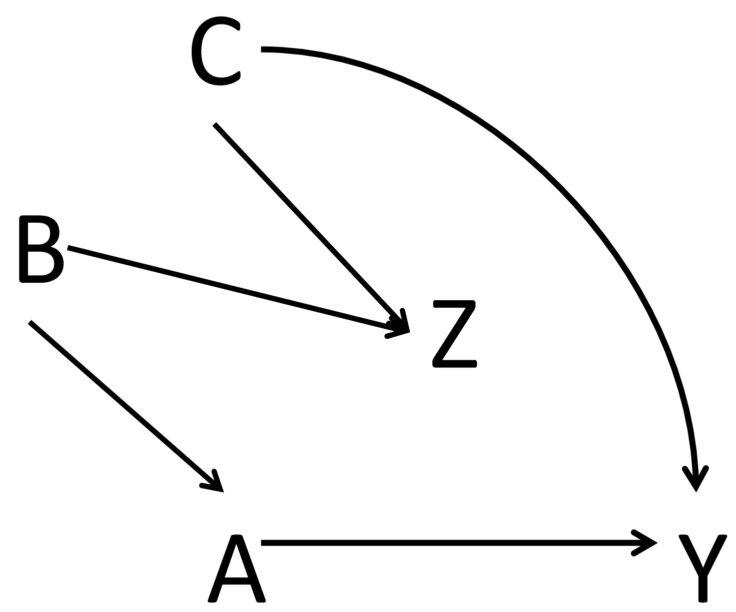


Collider example

d-connected (collider bias)

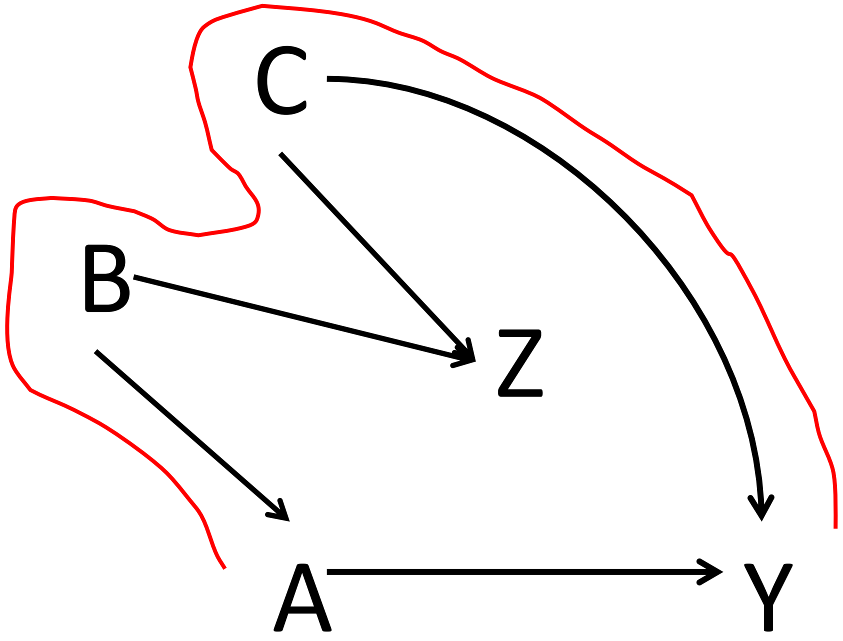


Example 1



How many paths between A and Y?
Are they causal/non-causal?
Are they open or blocked?
What set of variables is sufficient to control for confounding?

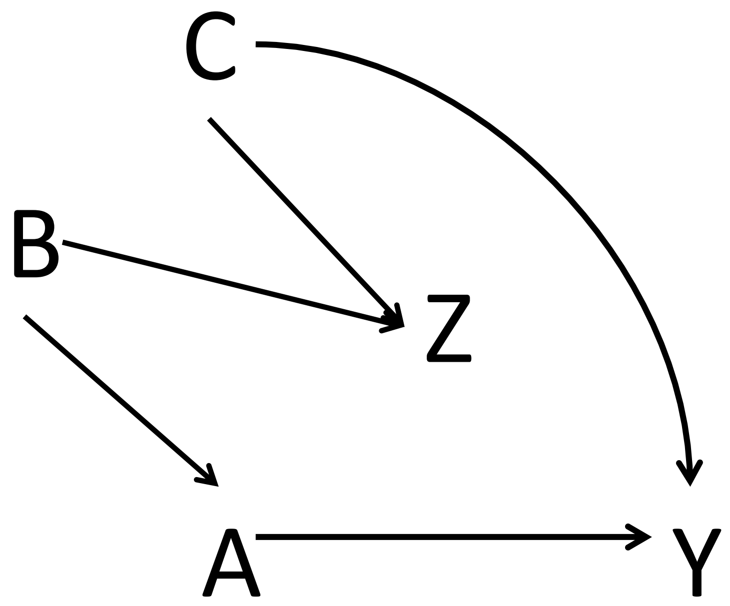
Example 1



Paths between A and Y

A → Y	Causal path (main)
A ← B → Z ← C → Y	Non-causal path, blocked (collider)

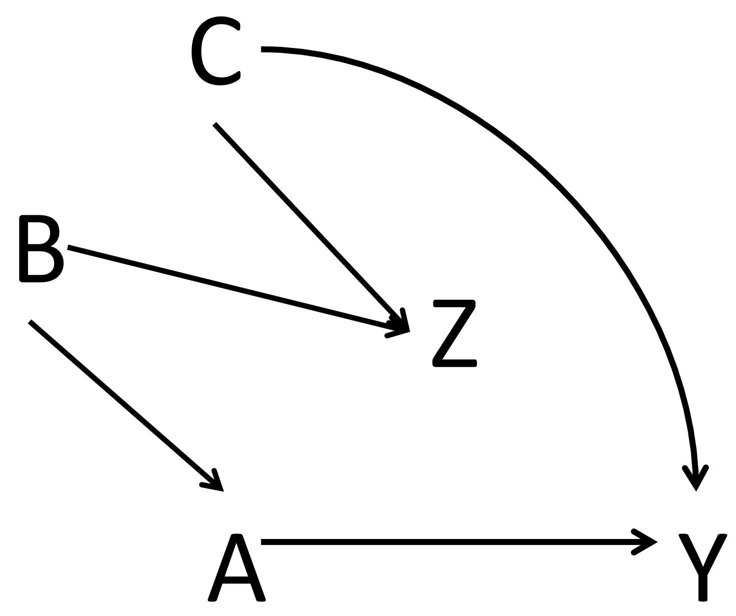
Example 1



Paths between A and Y

A → Y	Causal path (main)
A ← B → Z ← C → Y	Non-causal path, blocked (collider)

Example 1



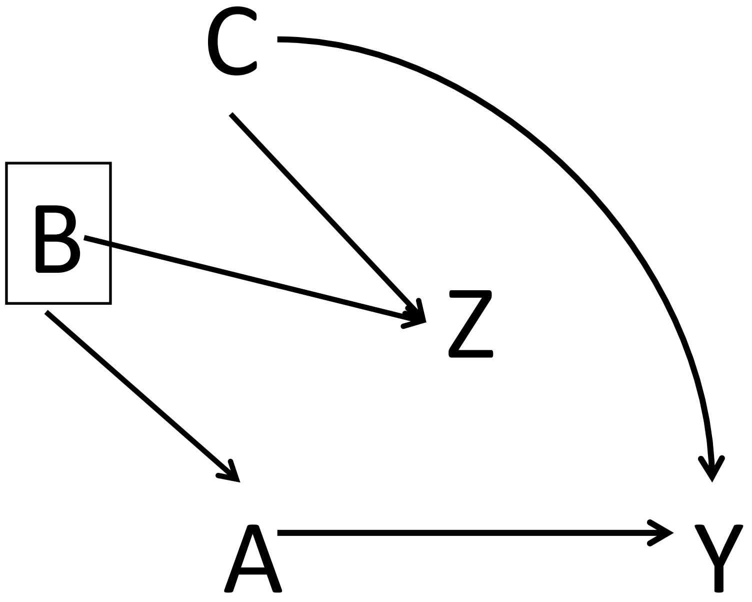
Paths between A and Y

A → Y	Causal path (main)
A ← B → Z ← C → Y	Non-causal path, blocked (collider)

Set of variables that is sufficient for adjustment

{ }

Example 1



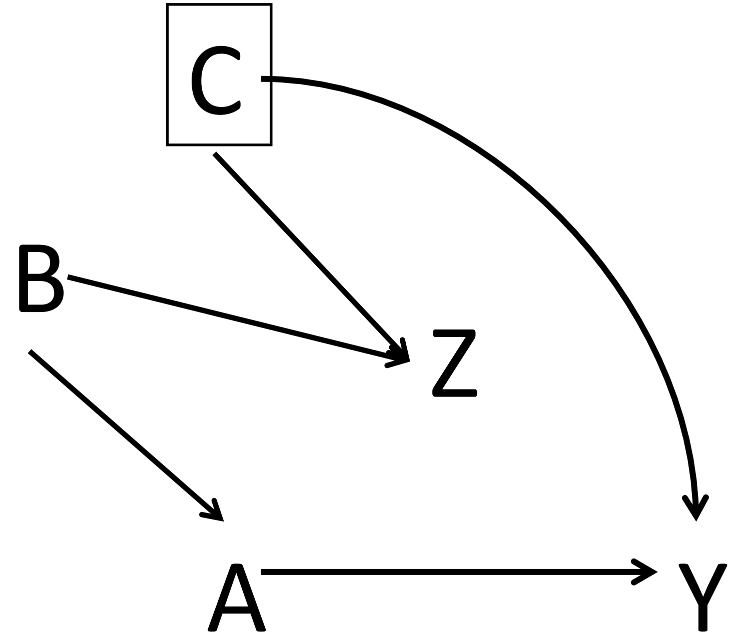
Paths between A and Y

A → Y	Causal path (main)
A ← B → Z ← C → Y	Non-causal path, blocked (collider)

Set of variables that is sufficient for adjustment

{ }
{B} OR {C} OR {B, C}

Example 1



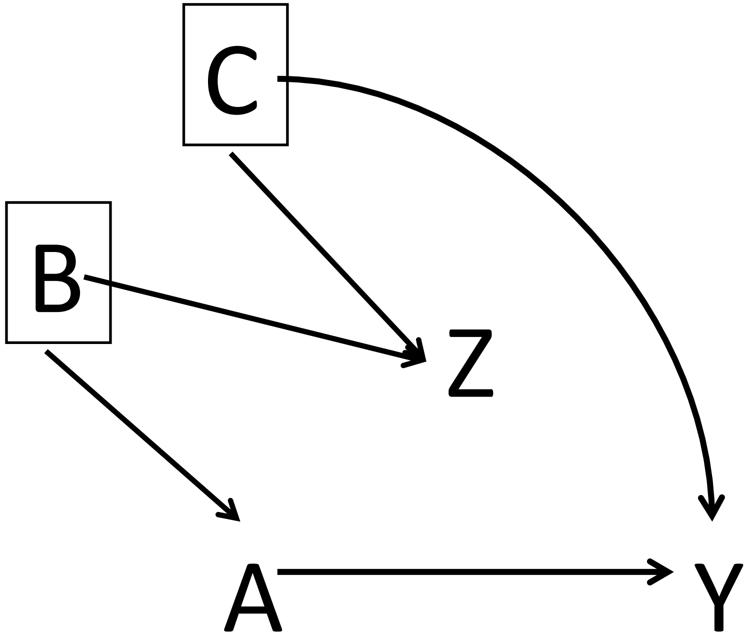
Paths between A and Y

A → Y	Causal path (main)
A ← B → Z ← C → Y	Non-causal path, blocked (collider)

Set of variables that is sufficient for adjustment

{ }
{B} OR {C} OR {B, C}

Example 1



Paths between A and Y

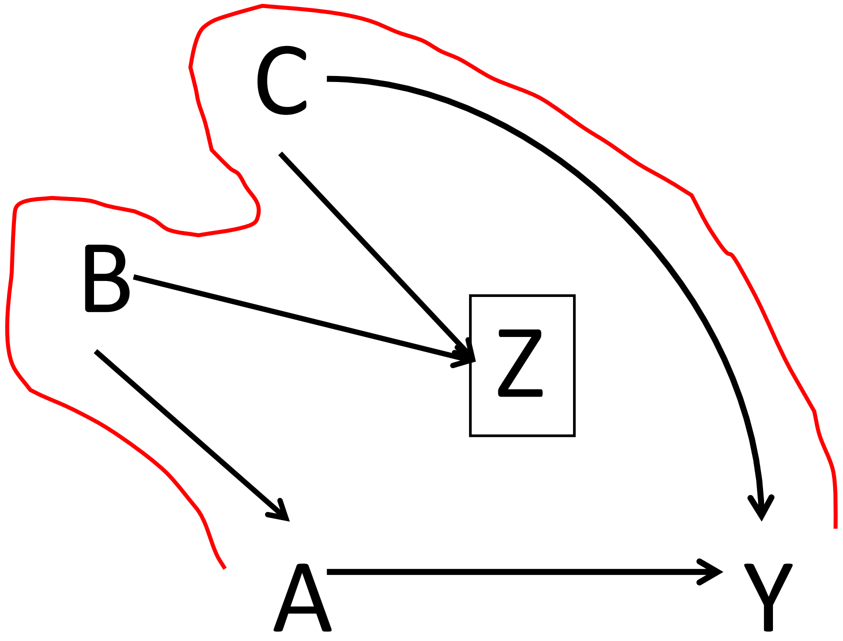
A → Y — Causal path (main)

A ← B → Z ← C → Y — Non-causal path, blocked (collider)

Set of variables that is sufficient for adjustment

{ }
{B} OR {C} OR {B, C}

Example 1



Paths between A and Y

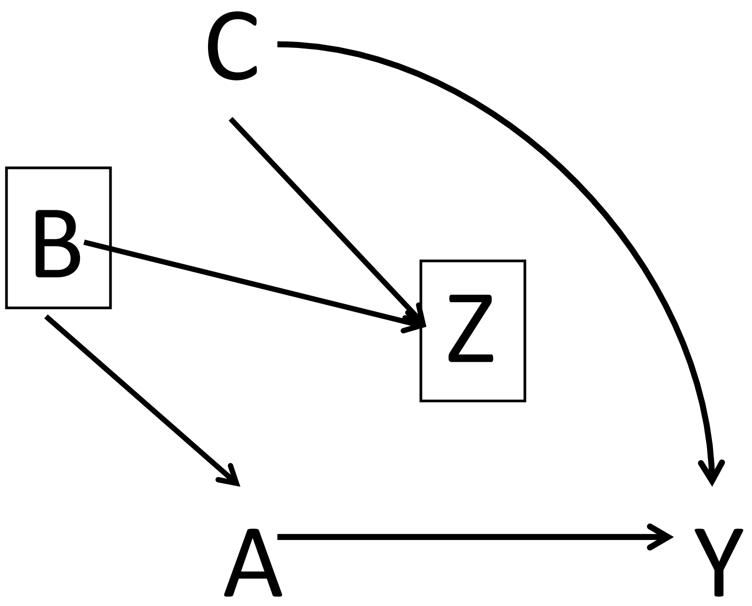
A → Y	Causal path (main)
A ← B → Z ← C → Y	Non-causal path, blocked (collider)

Set of variables that is sufficient for adjustment

{ }
{B} OR {C} OR {B, C}

If I control for Z, I must control for B, C, or both

Example 1



Paths between A and Y

A → Y Causal path (main)

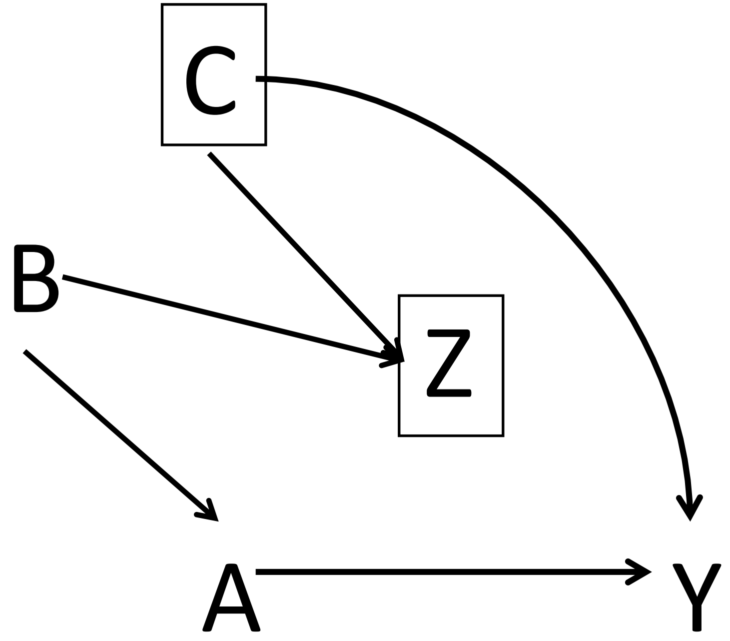
A ← B → Z ← C → Y Non-causal path, blocked (collider)

Set of variables that is sufficient for adjustment

{ }
{B} OR {C} OR {B, C}

If I control for Z, I must control for
B, C, or both

Example 1



Paths between A and Y

A → Y — Causal path (main)

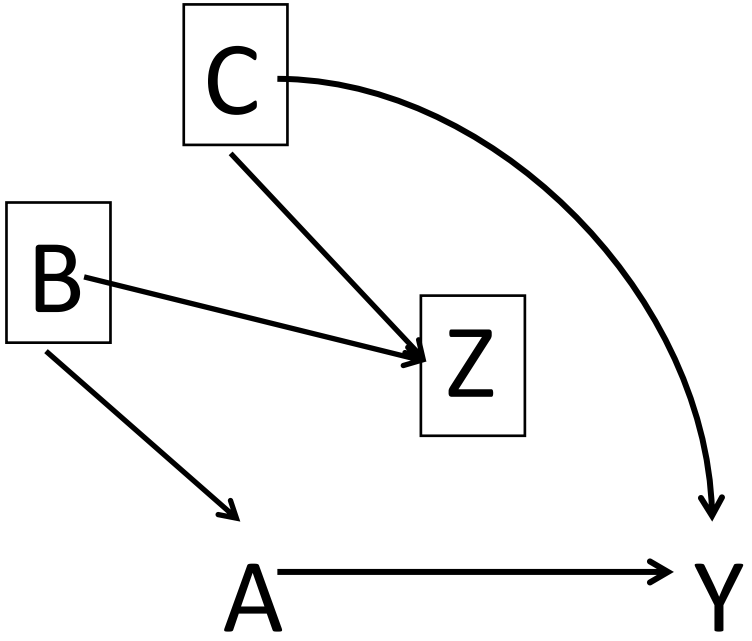
A ← B → Z ← C → Y — Non-causal path, blocked (collider)

Set of variables that is sufficient for adjustment

{ }
{B} OR {C} OR {B, C}

If I control for Z, I must control for B, C, or both

Example 1



Paths between A and Y

A → Y — Causal path (main)

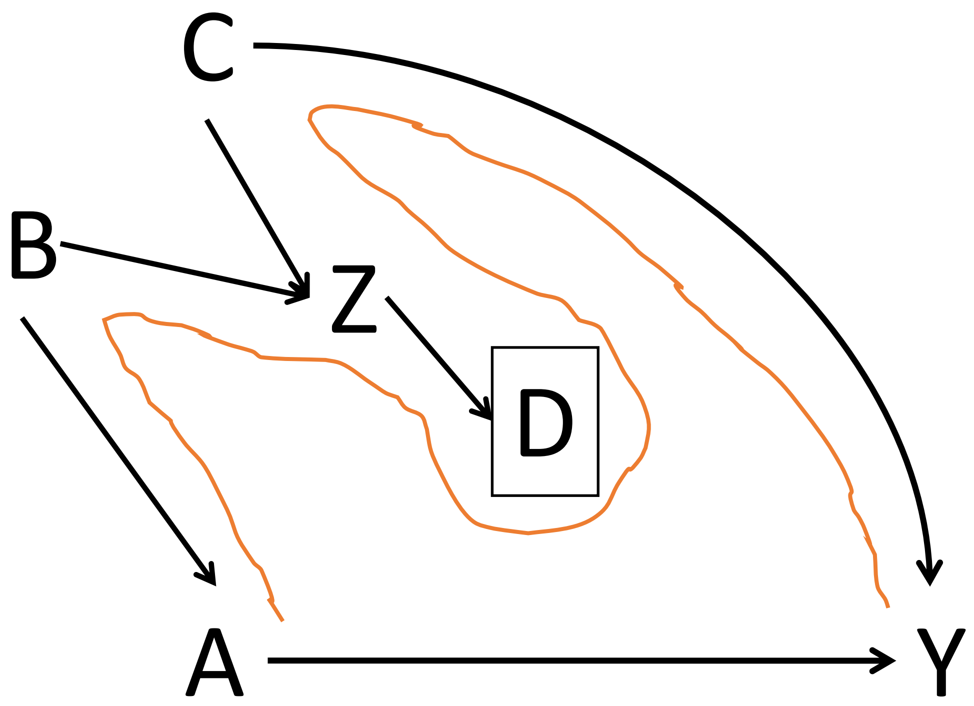
A ← B → Z ← C → Y — Non-causal path, blocked (collider)

Set of variables that is sufficient for adjustment

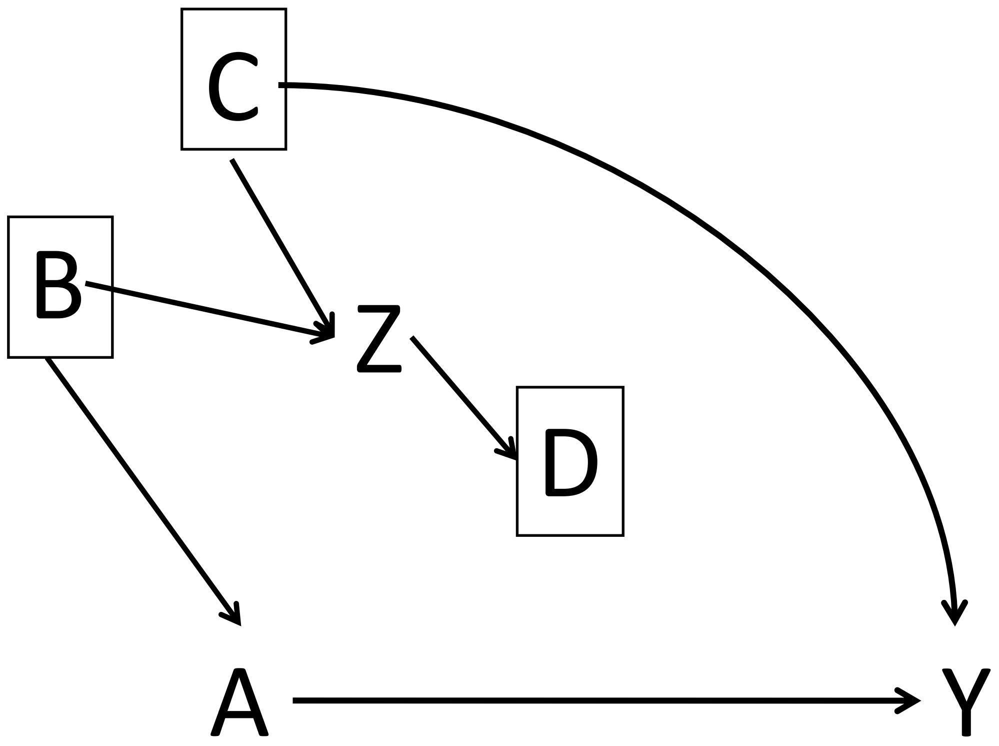
{ }
{B} OR {C} OR {B, C}

If I control for Z, I must control for B, C, or both

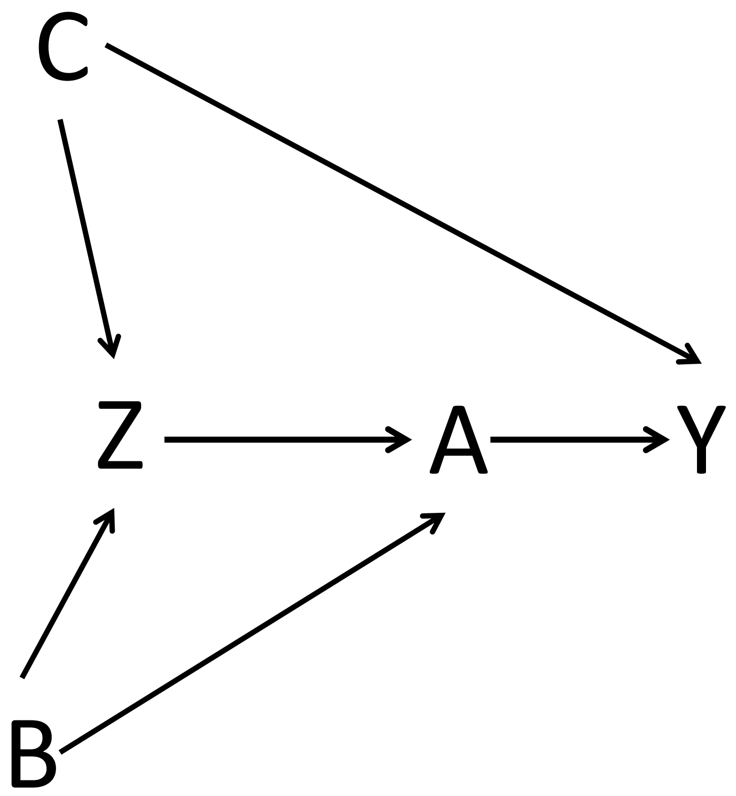
Example 1.1



Example 1.1

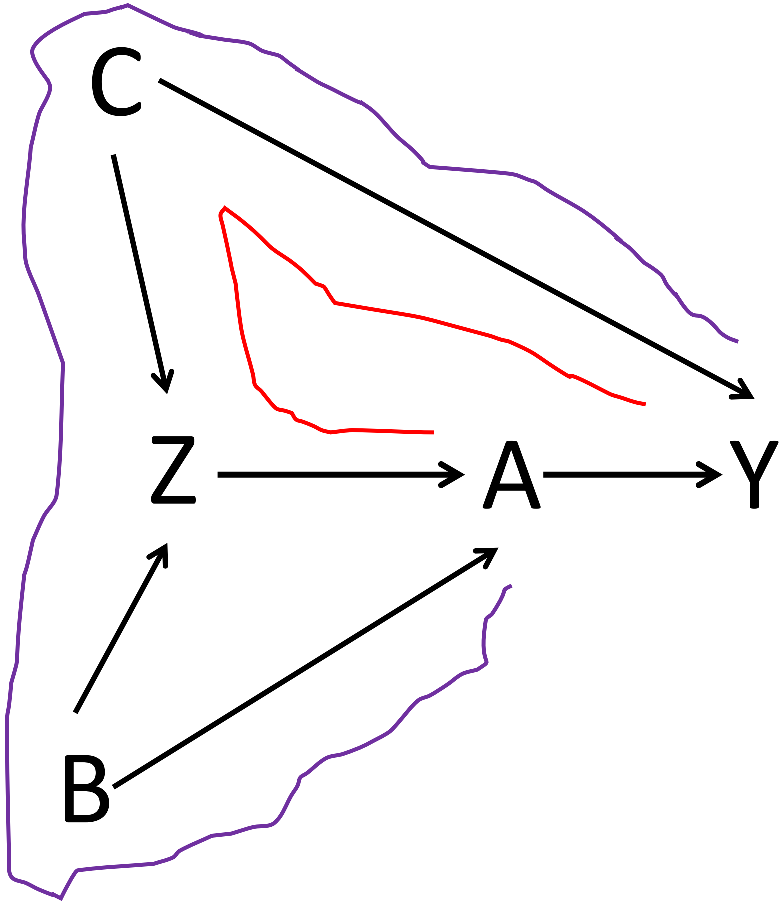


Example 2



How many paths between A and Y?
Are they causal/non-causal?
Are they open or blocked?
What set of variables is sufficient to control for confounding?

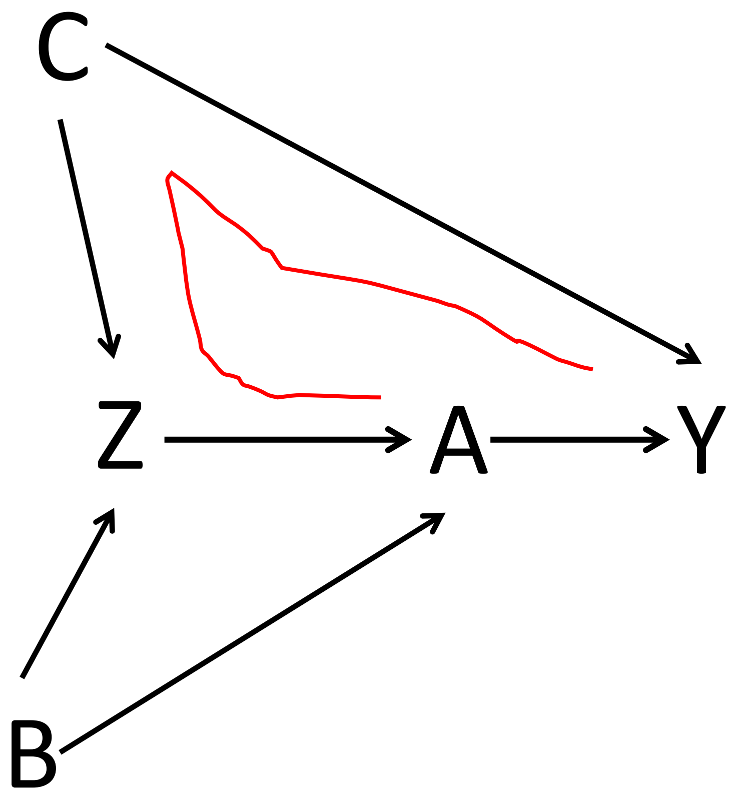
Example 2



Paths between A and Y

A → Y	Causal path (main)
A ← Z ← C → Y	Non-causal path, open
A ← B → Z ← C → Y	Non-causal path, blocked

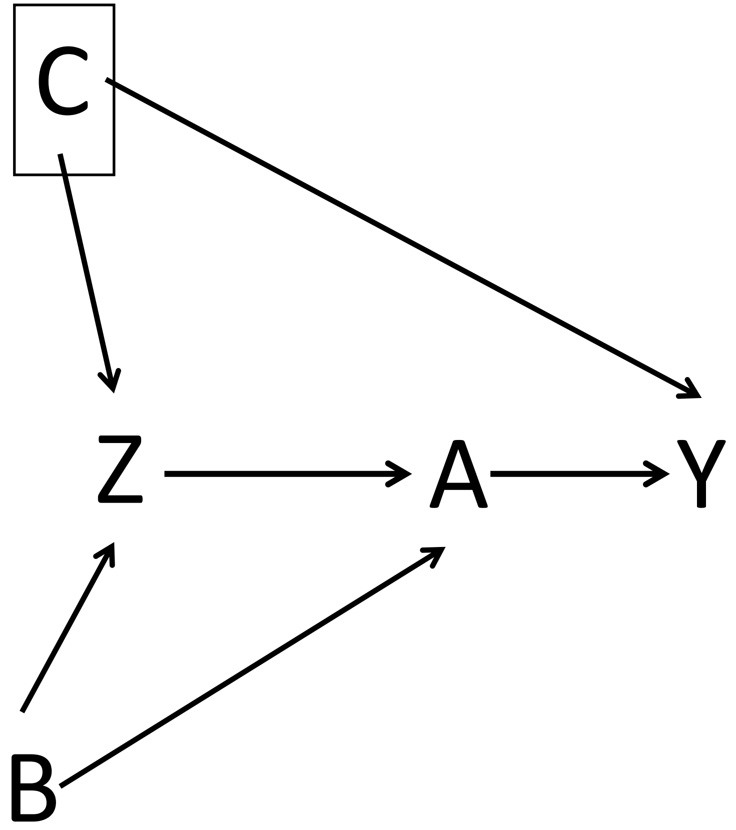
Example 2



Paths between A and Y

A → Y	Causal path (main)
A ← Z ← C → Y	Non-causal path, open
A ← B → Z ← C → Y	Non-causal path, blocked

Example 2



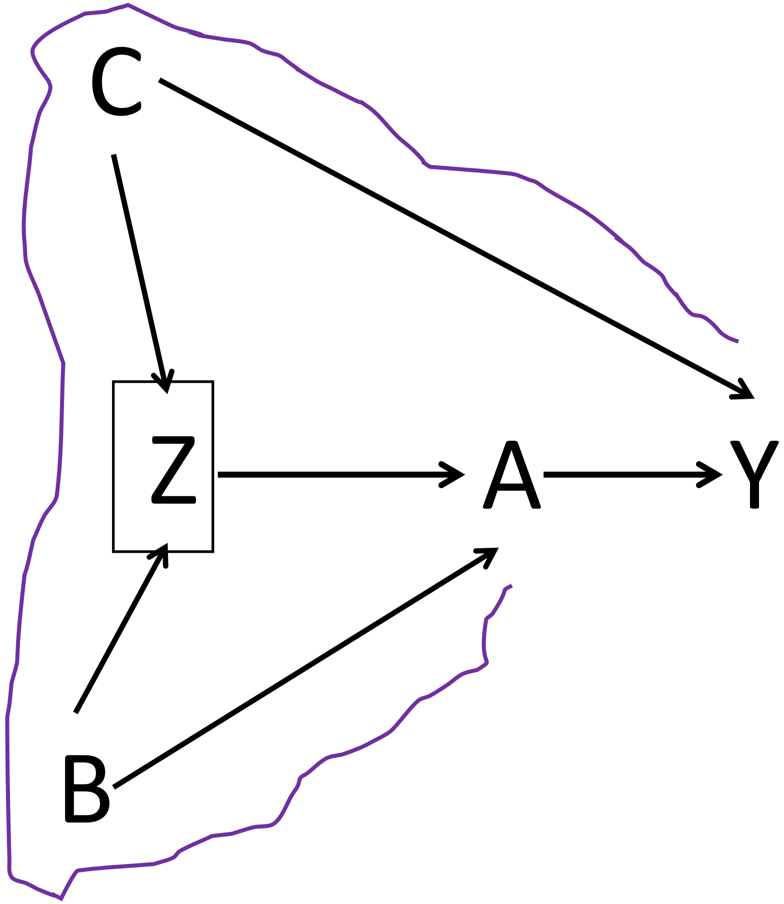
Paths between A and Y

A → Y	Causal path (main)
A ← Z ← C → Y	Non-causal path, open
A ← B → Z ← C → Y	Non-causal path, blocked

Set of variables that is sufficient for adjustment

{C}

Example 2



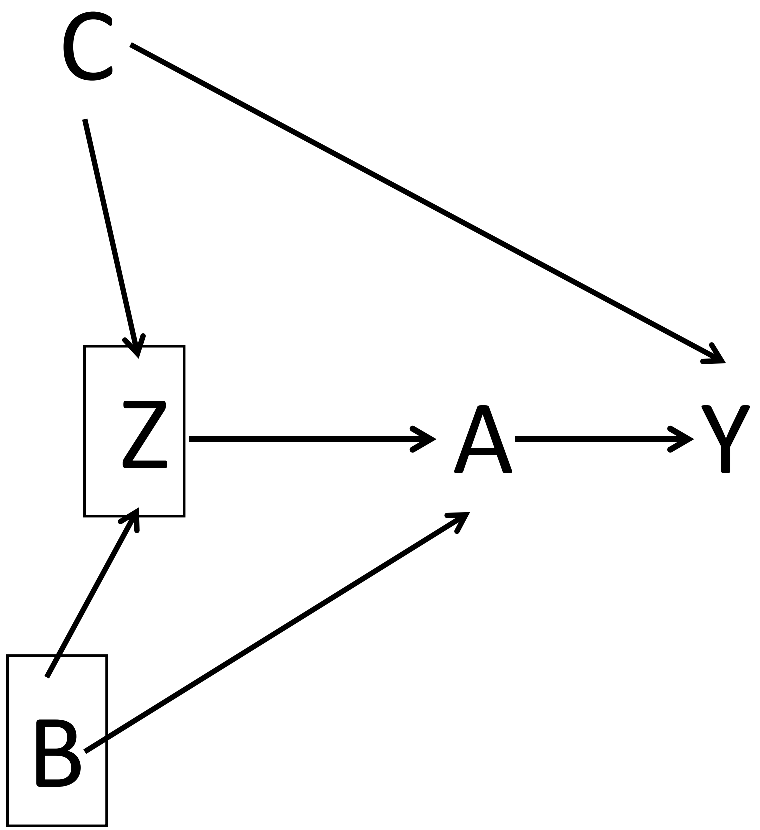
Paths between A and Y

A → Y	Causal path (main)
A ← Z ← C → Y	Non-causal path, open
A ← B → Z ← C → Y	Non-causal path, blocked

Set of variables that is sufficient for adjustment

{C}
{Z,B} OR {Z,C}

Example 2



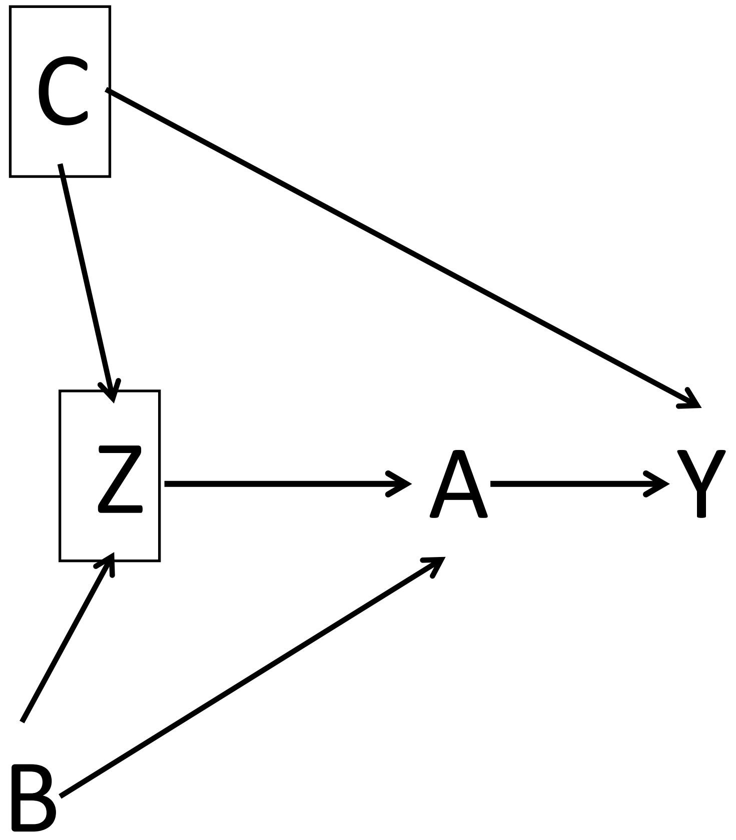
Paths between A and Y

A → Y	Causal path (main)
A ← Z ← C → Y	Non-causal path, open
A ← B → Z ← C → Y	Non-causal path, blocked

Set of variables that is sufficient for adjustment

{C}
{Z,B} OR {Z,C}

Example 2



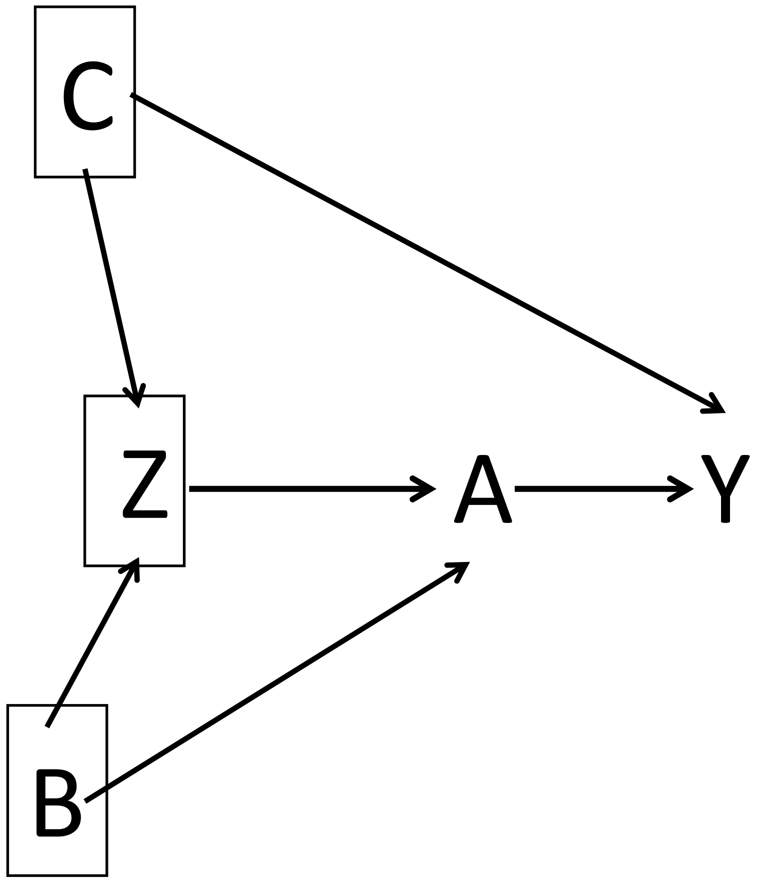
Paths between A and Y

A → Y	Causal path (main)
A ← Z ← C → Y	Non-causal path, open
A ← B → Z ← C → Y	Non-causal path, blocked

Set of variables that is sufficient for adjustment

{C}
{Z,B} OR {Z,C}

Example 2



Paths between A and Y

A → Y	Causal path (main)
A ← Z ← C → Y	Non-causal path, open
A ← B → Z ← C → Y	Non-causal path, blocked

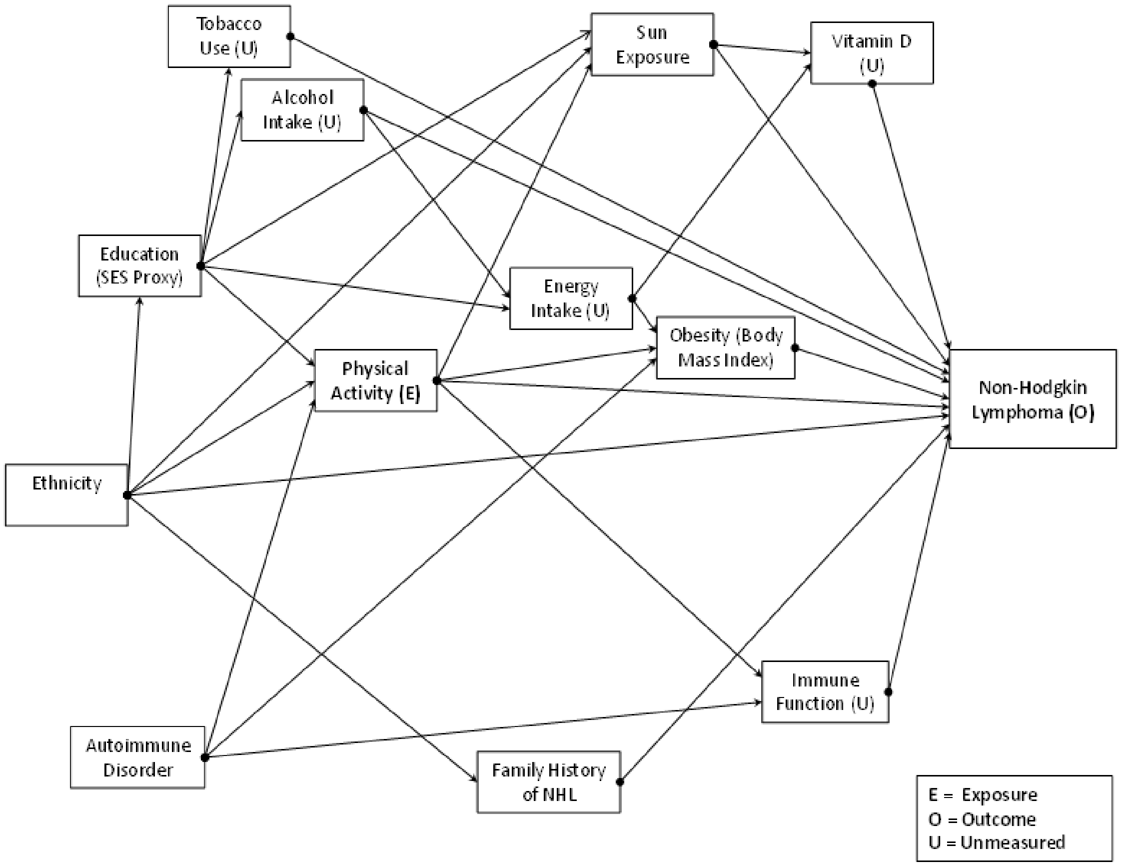
Set of variables that is sufficient for adjustment

{C}
{Z,B} OR {Z,C}
{Z, B, C}

If controls for Z, also must control for B, C, or both

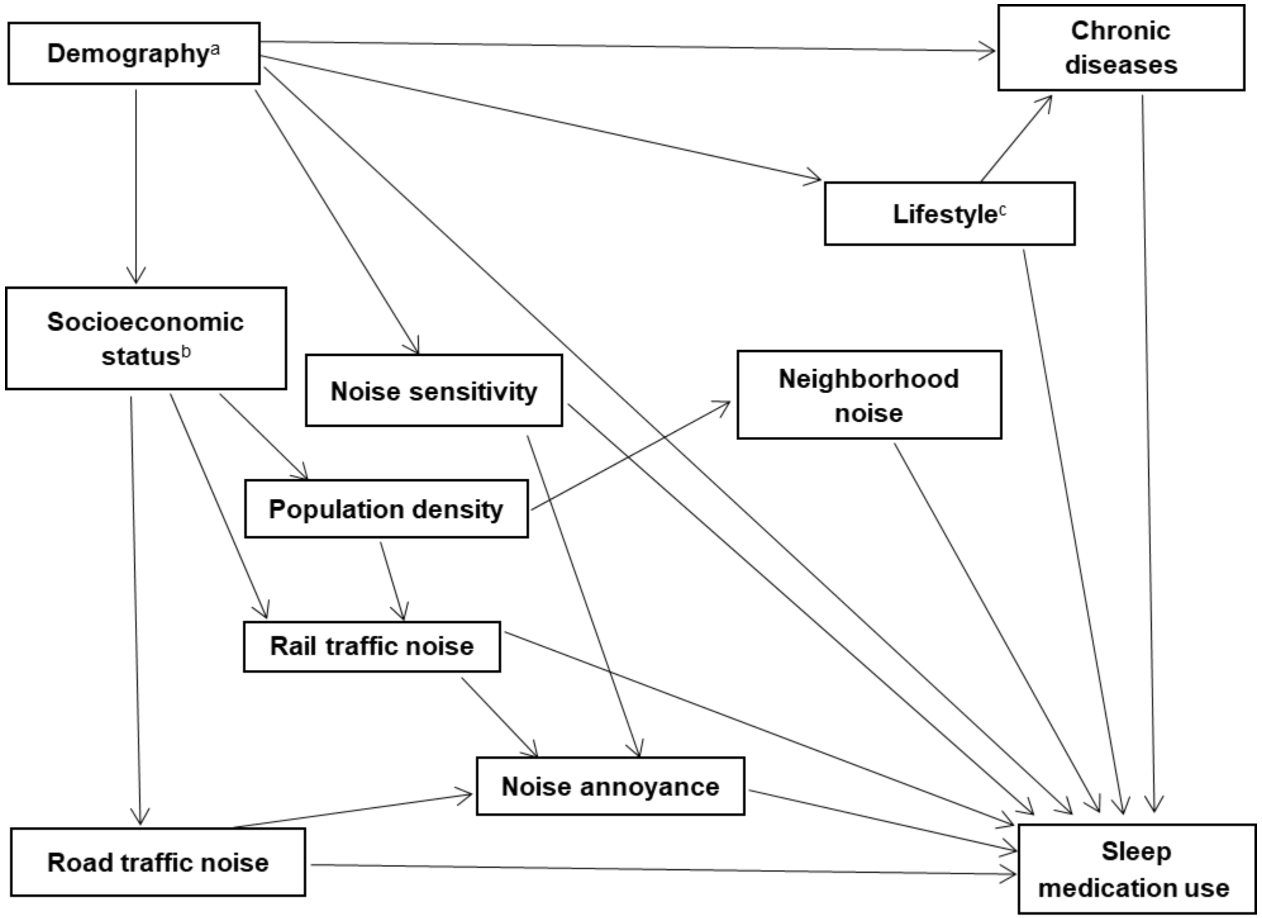
Questions

Example 1. DAG with unobserved confounders



Tennant PWG, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50(2):620-632. doi:10.1093/ije/dyaa213
Boyle T, et al. Lifetime physical activity and the risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev*. 2015;24(5):873-877. doi:10.1158/1055-9965.EPI-14-1303

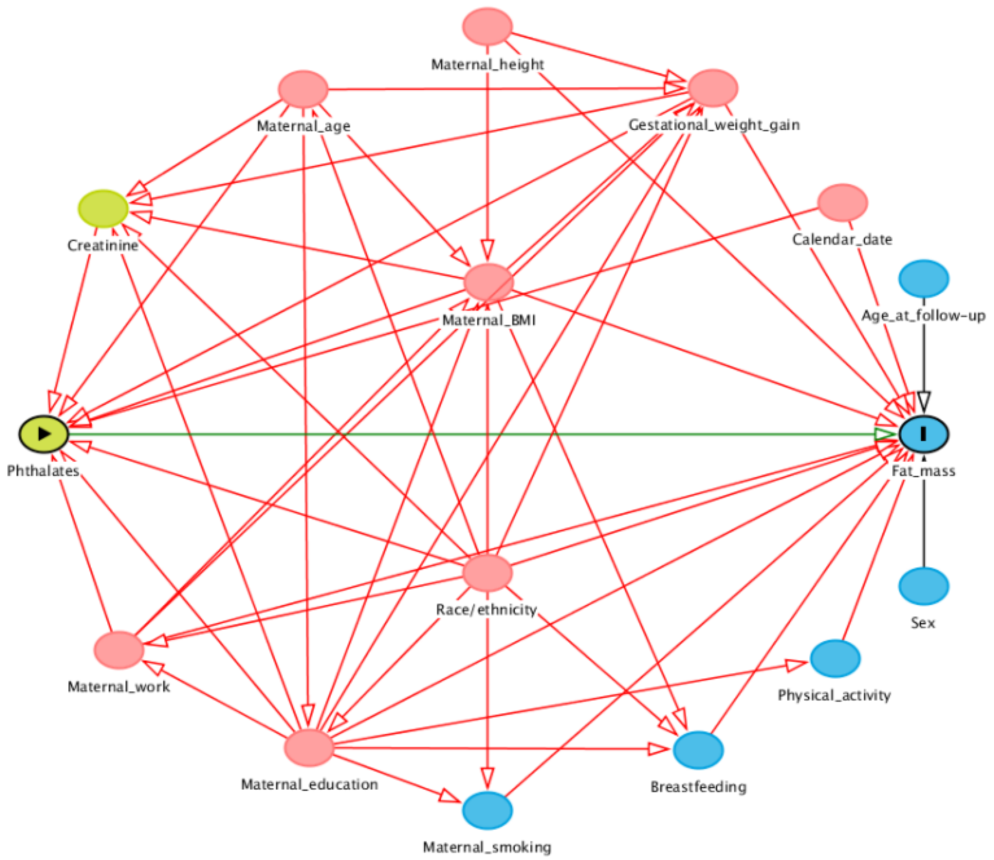
Example 2. DAG with super nodes



Tennant PWG, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50(2):620-632. doi:10.1093/ije/dyaa213

Evandt J, et al. Road traffic noise and registry based use of sleep medication. Environ Health. 2017;16(1):110. DOI: 10.1186/s12940-017-0330-5

Example 3. DAG where the flow of the arcs are inconsistent



Tennant PWG, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50(2):620-632. doi:10.1093/ije/dyaa213
Buckley JP, et al. Prenatal Phthalate Exposures and Childhood Fat Mass in a New York City Cohort. Environ Health Perspect. 2016; 124(4): 507–513. DOI: 10.1289/ehp.1509788

Checklist of recommendations for studies using DAGs

Section	Recommendation number	Recommendation details	Page
Introduction			
	1	The focal relationship(s) and estimand(s) of interest are stated in the study aims	
Methods			
	2	DAGs for all focal relationships and estimands of interest are provided	
	3	DAGs include all relevant variables, including those where direct measurements are unavailable	
	4	DAGs are arranged so that all constituent arcs flow in the same direction	
	5	Missing arcs have been carefully considered. Optionally, these are justified with theory and/or evidence.	
	6	DAG-implied adjustment set(s) for all estimand(s) of interest are clearly stated, including any unobserved confounders	
	8a	Alternative adjustment set(s) are clearly described and justified	
	8b	Optionally, the consistency of all DAGs with the observed data has been explored. Subsequently modified DAGs are reported separately	
Results			
	7a	Estimate(s) from unmodified DAG-implied adjustment set(s) - or the nearest approximation thereof - are provided	
	7b	Optionally, the impact of unobserved confounders has been estimated, and bias-adjusted estimates are reported	
	8c	Estimates from alternative adjustment set(s) are reported separately to those obtained from DAG-implied adjustment sets	

Tennant PWG, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50(2):620-632. doi:10.1093/ije/dyaa213

Questions

DAGitty

www.dagitty.net

Literature

References:

- Pearl J. Causal diagrams for experimental research. Biometrika, 82(4):669-710, 1995.
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology (Cambridge, Mass). 1999;10(1):37-48.
- Staplin et. al, 2017. Use of causal diagrams to inform the design and interpretation of observational studies: An example from the Study of Heart and Renal Protection (SHARP). Clin J Am Soc Nephrol. 2017 Mar 7;12(3):546-552.
- Tennant PWG, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50(2):620-632. doi:10.1093/ije/dyaa213

Recommended readings:

- Elwert, Felix (2013). “Graphical causal models". In: Handbook of causal analysis for social research. Springer, pp. 245–273. Available: https://www.researchgate.net/publication/278717528_Graphical_Causal_Models
- Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC. Available: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2020/02/ci_hernanrobins_21feb20.pdf
- Check the ‘Introduction Literature on DAGs’ section of the ‘Learn’ option of DAGitty website for more introduction readings: http://www.dagitty.net/learn/index.html

Thank you!

Ione Avila-Palencia, PhD, MPH

i.avila-palencia@qub.ac.uk

Hüseyin Küçükali, MD, PhD

h.kucukali@qub.ac.uk

This work was supported by **UK Research and Innovation** [ES/V016075/1]

Economic and Social Research Council

Healthy Ageing Challenge Social, Behavioural and Design Research