LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Confounded about confounding?

How 'modern' causal thinking brings fresh clarity to old concerns

Bianca De Stavola and Rhian Daniel

Thursday 6th October, 2011 · 12.30pm

Outline

Outline

Introduction

Background

- Recognition that our statistical language and analytical tools were designed to study association – **not** causation
- Main limitation: likely presence of confounding bias.
- Yet, public health implications of results from associational models are (nearly) always mentioned.
- Agreement that our teaching (and practice) of how to deal with confounding bias is inadequate.
- Skepticism regarding whether anything can be done.
- *Can modern thinking on causal inference be helpful?*

Introduction

Aims

1. Review traditional methods and highlight when they might be inadequate
2. Introduce causal language and demonstrate it is necessary if we wish to make causal statements
3. Define some essential tools (DAGs) and illustrate their use
4. Give a brief overview of the choice of existing and **new** statistical methods
5. Broaden the debate

Traditional tools

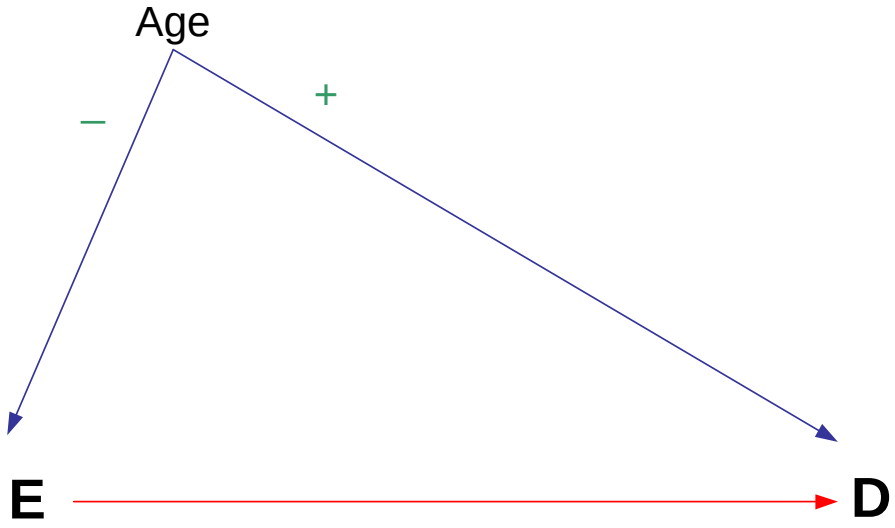
An illustrative example (simulated data)

Question:

Does eating dark green vegetables protect against stomach cancer?

Simulated data:

- Fixed term cohort study: 752 cancers out of 10,000 female nurses (aged 30-70y)
- Exposure: usual intake at baseline
- Potential confounders: age, . . .



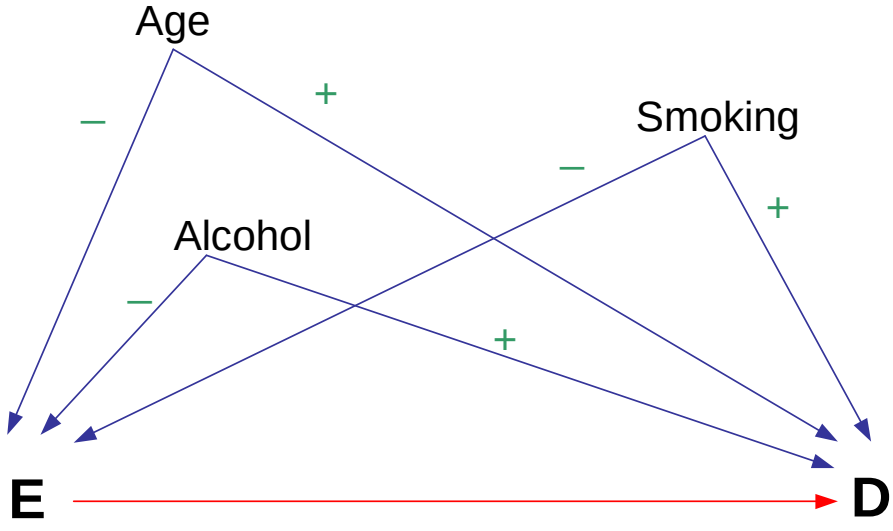
Traditional tools

An illustrative example (simulated data)

Question:

Does eating dark green vegetables protect against stomach cancer?

- Fixed term cohort study: 752 cancers out of 10,000 female nurses (aged 30-70y)
- Exposure: usual intake at baseline
- Potential confounders: age, smoking, alcohol intake



Results

Results

Table: Odds ratio (OR) of dark green vegetable intake (x 1SD)

Adjusted for:	**OR**	**95%CI**
Age	1.04	0.97, 1.11
Age + alc	1.52	1.40, 1.66
Age + smok	1.03	0.96, 1.11
Age+alc+smok	1.39	1.27, 1.52

Results

Results

Table: Odds ratio (OR) of dark green vegetable intake (x 1SD)

Adjusted for:	**OR**	**95%CI**
Age	1.04	0.97, 1.11
Age + alc	1.52	1.40, 1.66
Age + smok	1.03	0.96, 1.11
Age+alc+smok	1.39	1.27, 1.52

- Which OR to report (assuming no interactions)?

Interpretation

An illustrative example (simulated data)

- How should we interpret the one we will choose?

Interpretation

An illustrative example (simulated data)

- How should we interpret the one we will choose?

  Say we adjust for all three

  a) *the expected effect of 1 unit increase in E, holding the confounders constant?*

Interpretation

An illustrative example (simulated data)

- How should we interpret the one we will choose?

  Say we adjust for all three

  a) *the expected effect of 1 unit increase in E, holding the confounders constant?*

  b) *Among those of a particular age, smoking & alcohol status:*

  b1) *the average percentage increase in cancer odds comparing those whose intake was $e+1$ versus those whose intake was e?*

Interpretation

An illustrative example (simulated data)

- How should we interpret the one we will choose?

  Say we adjust for all three

  a) *the expected effect of 1 unit increase in E, holding the confounders constant?*

  b) *Among those of a particular age, smoking & alcohol status:*

  b1) *the average percentage increase in cancer odds comparing those whose intake was $e+1$ versus those whose intake was e?*

  b2) *the expected percentage increase in cancer odds were everybody to add 1 unit to their dark green vegetable intake?*

Interpretation

An illustrative example (simulated data)

- How should we interpret the one we will choose?

  Say we adjust for all three

  a) *the expected effect of 1 unit increase in E, holding the confounders constant?*

  b) *Among those of a particular age, smoking & alcohol status:*
     - b1) *the average percentage increase in cancer odds comparing those whose intake was $e+1$ versus those whose intake was e?*
     - b2) *the expected percentage increase in cancer odds were everybody to add 1 unit to their dark green vegetable intake?*

  c) *the predicted percentage increase in cancer odds for an individual if s/he increased dark green vegetable intake by 1 unit, without changing age, smoking & alcohol status?*

Outline

A simple example

- 12 subjects each suffer a headache.
- Some take a potion; others don't.
- One hour later, we ask each of the 12 whether or not his/her headache has disappeared.

The observed data (1)

Here are the data:

	X (potion taken?)	Y (headache disappeared?)
Arianrhod	0	0
Blodeuwedd	1	0
Caswallawn	1	1
Dylan	0	0
Efnisien	0	1
Gwydion	1	0
Hafgan	1	0
Lleu	0	0
Matholwch	0	1
Pwyll	0	0
Rhiannon	0	1
Teyrnon	1	1

The observed data (2)

Here are the data:

	X (potion taken?)	Y (headache disappeared?)
Arianrhod	0	0
Blodeuwedd	1	0
Caswallawn	1	1
Dylan	0	0
Efnisien	0	1
Gwydion	1	0
Hafgan	1	0
Lleu	0	0
Matholwch	0	1
Pwyll	0	0
Rhiannon	0	1
Teyrnon	1	1

- Caswallawn took the potion, and his headache disappeared.
- Did the potion cause his headache to disappear?
- We don't know.
- To answer this, we need to know what would have happened had he not taken the potion.

Counterfactuals and potential outcomes

- X is the treatment: whether or not a potion was taken.
- Y is the outcome: whether or not the headache disappeared.
- Write Y^0 and Y^1 to represent the *potential outcomes* under both treatments.
- Y^0 is the outcome which would have been seen had the potion NOT been taken.
- Y^1 is the outcome which would have been seen had the potion been taken.
- One of these is observed: if $X = 0$, Y^0 is observed; if $X = 1$, Y^1 is observed.
- The other is *counterfactual*.
- Suppose that we can observe the unobservable. . .

The ideal data (1)

The 'ideal' data:

	Y^0	Y^1
Arianrhod	0	0
Blodeuwedd	1	0
Caswallawn	0	1
Dylan	0	0
Efnisien	1	1
Gwydion	0	0
Hafgan	0	0
Lleu	0	0
Matholwch	1	0
Pwyll	0	0
Rhiannon	1	1
Teyrnon	0	1

- For Caswallawn, the potion did have a causal effect.
- He did take it, and his headache disappeared; but had he not taken it, his headache would not have disappeared.
- Thus the potion had a causal effect on his headache.
- What about Gwydion?
- and Rhiannon?
- and Matholwch?

The ideal data (2)

The 'ideal' data:

	Y^0	Y^1	Causal effect?
Arianrhod	0	0	No
Blodeuwedd	1	0	Yes, harmful
Caswallawn	0	1	Yes, protective
Dylan	0	0	No
Efnisien	1	1	No
Gwydion	0	0	No
Hafgan	0	0	No
Lleu	0	0	No
Matholwch	1	0	Yes, harmful
Pwyll	0	0	No
Rhiannon	1	1	No
Teyrnon	0	1	Yes, protective

- An individual-level causal effect is defined for each subject and is given by

$$Y^1 - Y^0$$

- These need not all be the same.

The fundamental problem of causal inference

Back to reality...

	Y^0	Y^1	X	Y
Arianrhod	0	?	0	0
Blodeuwedd	?	0	1	0
Caswallawn	?	1	1	1
Dylan	0	?	0	0
Efnisien	1	?	0	1
Gwydion	?	0	1	0
Hafgan	?	0	1	0
Lleu	0	?	0	0
Matholwch	1	?	0	1
Pwyll	0	?	0	0
Rhiannon	1	?	0	1
Teyrnon	?	1	1	1

- In reality, we never observe both Y^0 and Y^1 on the same individual.
- Sometimes called the fundamental problem of causal inference.
- It is therefore over-ambitious to try to infer anything about individual-level causal effects.

Population-level causal effects (1)

- A less ambitious goal is to focus on the population-level or average causal effect:

$$E\left(Y^1\right) - E\left(Y^0\right)$$

or, since Y is binary,

$$P\left(Y^1 = 1\right) - P\left(Y^0 = 1\right)$$

- Let's return to the 'ideal' data...

Population-level causal effects (2)

	Y^0	Y^1	Causal effect?
Arianrhod	0	0	No
Blodeuwedd	1	0	Yes, harmful
Caswallawn	0	1	Yes, protective
Dylan	0	0	No
Efnisien	1	1	No
Gwydion	0	0	No
Hafgan	0	0	No
Lleu	0	0	No
Matholwch	1	0	Yes, harmful
Pwyll	0	0	No
Rhiannon	1	1	No
Teyrnon	0	1	Yes, protective

$$P\left(Y^0 = 1\right) = \frac{4}{12}$$

$$P\left(Y^1 = 1\right) = \frac{4}{12}$$

$$P\left(Y^1 = 1\right) - P\left(Y^0 = 1\right) = 0$$

i.e. no causal effect at the population level.

Population-level causal effects (3)

- In reality, we don't know Y^1 for every subject, so we can't simply estimate $P\left(Y^1 = 1\right)$ as the proportion of all subjects with $Y^1 = 1$.
- Likewise, we can't simply estimate $P\left(Y^0 = 1\right)$ as the proportion of all subjects with $Y^0 = 1$.
- Thus we can't easily estimate $P\left(Y^1 = 1\right) - P\left(Y^0 = 1\right)$ for the same reason that we can't estimate $Y^1 - Y^0$.
- Causal inference is all about choosing quantities from the observed data (i.e. involving X, Y and other observed variables) that represent reasonable substitutes for hypothetical quantities such as $P\left(Y^1 = 1\right) - P\left(Y^0 = 1\right)$, which involve unobservable counterfactuals.

When does association = causation? (1)

- What might be a good substitute for $P\left(Y^1 = 1\right)$?
- What about $P\left(Y = 1 \,|X = 1\right)$?
- This is the proportion whose headache disappeared among those who actually took the potion.
- Is this the same as $P\left(Y^1 = 1\right)$?
- Only if those who took the potion are **exchangeable** with those who didn't. [Mathematically: $\left\{Y^0, Y^1\right\} \perp\!\!\!\perp X$]
- This would be the case if the choice to take the potion was made at random.
- This is why ideal randomised experiments are the gold standard for inferring causal effects.

When does association = causation? (2)

	Y^0	Y^1	X	Y
Arianrhod	0	?	0	0
Blodeuwedd	?	0	1	0
Caswallawn	?	1	1	1
Dylan	0	?	0	0
Efnisien	1	?	0	1
Gwydion	?	0	1	0
Hafgan	?	0	1	0
Lleu	0	?	0	0
Matholwch	1	?	0	1
Pwyll	0	?	0	0
Rhiannon	1	?	0	1
Teyrnon	?	1	1	1

$$P(Y = 1 \,|X = 1) = \frac{2}{5}$$

$$P(Y = 1 \,|X = 0) = \frac{3}{7}$$

$$P(Y = 1 \,|X = 1) - P(Y = 1 \,|X = 0) = -\frac{1}{35}$$

If we assumed that association = causation, we would conclude that the potion was, on average, slightly harmful.

What's going on here?

	Y^0	Y^1	X	Y
Arianrhod	0	0	0	0
Blodeuwedd	1	0	1	0
Caswallawn	0	1	1	1
Dylan	0	0	0	0
Efnisien	1	1	0	1
Gwydion	0	0	1	0
Hafgan	0	0	1	0
Lleu	0	0	0	0
Matholwch	1	0	0	1
Pwyll	0	0	0	0
Rhiannon	1	1	0	1
Teyrnon	0	1	1	1

- The subjects with the more severe headaches are more likely to take the potion.
- So association $\neq$ causation.

Taking severity into account

- Suppose we asked each of the 12 subjects at the beginning of the study: "is your headache severe?".
- Then, we might propose that, after taking severity into account, the decision as to whether or not to take the potion was effectively taken at random.
- Suppose Z denotes severity. Then, under this assumption, within strata of Z, the exposed and unexposed subjects are exchangeable.
- This is called **conditional exchangeability** (given Z).
  [Mathematically: $\{Y^0, Y^1\} \perp\!\!\!\perp X \,|\, Z$]
- Under conditional exchangeability given Z, association = causation within strata of Z.
- Let's return to the data and look for an association between X and Y within strata of Z.

Stratifying on severity

	Y^0	Y^1	X	Y	Z
Arianrhod	0	0	0	0	1
Blodeuwedd	1	0	1	0	0
Caswallawn	0	1	1	1	0
Dylan	0	0	0	0	1
Efnisien	1	1	0	1	0
Gwydion	0	0	1	0	1
Hafgan	0	0	1	0	1
Lleu	0	0	0	0	0
Matholwch	1	0	0	1	1
Pwyll	0	0	0	0	0
Rhiannon	1	1	0	1	0
Teyrnon	0	1	1	1	1

In the stratum $Z = 0$:

$$P(Y = 1 | X = 1) = \frac{1}{2}$$

$$P(Y = 1 | X = 0) = \frac{2}{4}$$

In the stratum $Z = 1$:

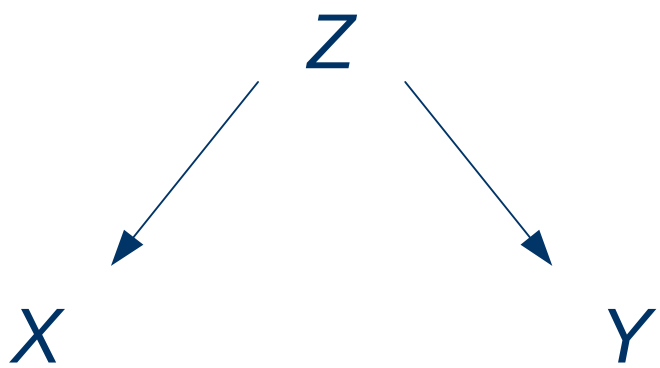
$$P(Y = 1 | X = 1) = \frac{1}{3}$$

$$P(Y = 1 | X = 0) = \frac{1}{3}$$

i.e. within strata of Z we find no association between X and Y.

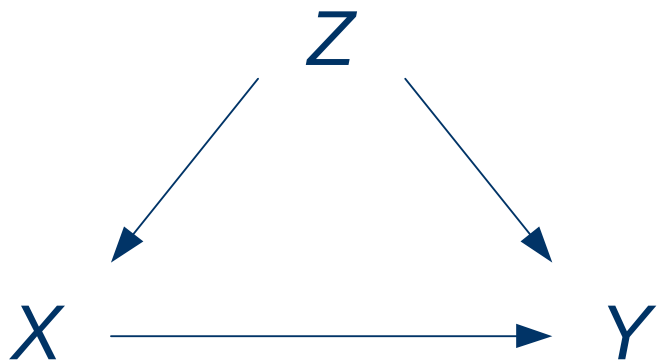
Summary so far (1)

- We have looked at a simple, artificial example, and defined what we mean by a causal effect.
- We have seen that, unless the exposed and unexposed groups are exchangeable, association is not causation.
- In our simple example, there was no (average) causal effect of X on Y.
- And yet, X and Y were associated, because of Z.



Summary so far (2)

- When we stratified on Z, we found no association between X and Y.
- So association = causation within strata of Z.
- This is because exposed and unexposed subjects were conditionally exchangeable given Z.
- More generally, when there is a causal effect of X on Y, but also a non-causal association via Z, the causal effect will be estimated with bias unless we stratify on Z.



Summary so far (3)

- Conditional exchangeability is the key criterion that allows us to make causal statements using observational data.
- Thus we need to identify, if possible, a set of variables Z_1, Z_2, . . . , such that conditional exchangeability holds given these.
- In real life, there may be many many candidate Z-variables.
- These may be causally inter-related in a very complex way.
- Deciding whether or not the exposed and unexposed are conditionally exchangeable given Z_1, Z_2, . . . requires detailed background subject-matter knowledge.
- Causal diagrams can help us to use this knowledge to determine whether or not conditional exchangeability holds.

Outline

Outline

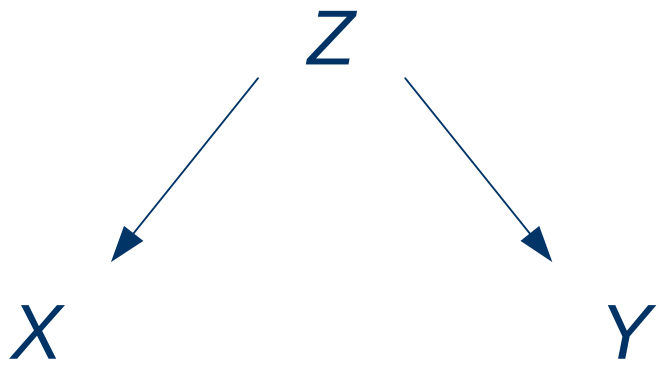
How can two variables be associated in the population? (1)

- Two variables X and Y will be associated in the population if X causes Y.

How can two variables be associated in the population? (2)

- X and Y will also be associated if Y causes X.

How can two variables be associated in the population? (3)



- Finally, X and Y will also be associated if there is some Z that causes both X and Y.

How can two variables be associated in the population? (4)

- X and Y cannot be associated in the population for any other reason.
- If X and Y are associated in the population then at least one of the above must be true.

What do we mean by associated 'in the population'?

- In statistical terminology, X and Y being associated 'in the population' means that they are marginally associated.
- If X and Y are marginally associated, then, for a particular subject, knowing the value of X gives us some information about the likely value of Y and vice versa.
- Suppose, for simplicity, that X and Y are both binary. If X and Y are marginally associated then

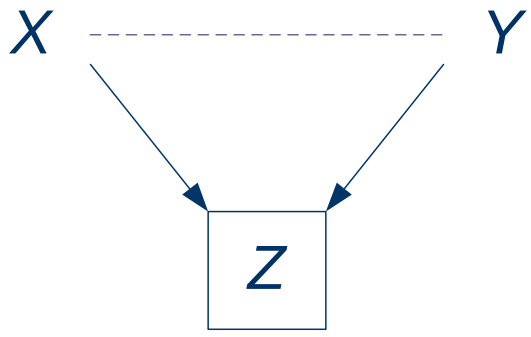
$$P(X = 1 \,|\, Y = 1) \neq P(X = 1 \,|\, Y = 0)$$

and

$$P(Y = 1 \,|\, X = 1) \neq P(Y = 1 \,|\, X = 0)$$

- Next, we will talk about conditional association or association in a subpopulation.

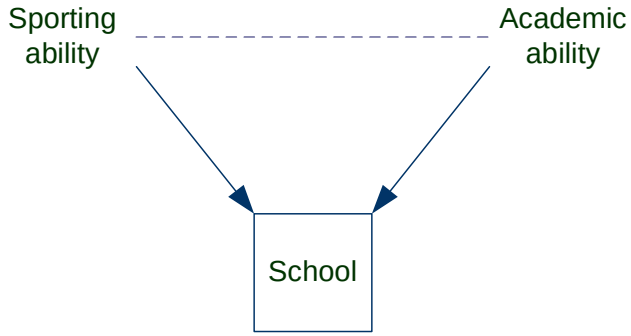
How can two variables be associated in a sub-population? (1)



- Suppose that Z is an effect of both X and Y.
- Then X and Y will be associated within strata of Z, even if they are independent in the population.
- The box around Z denotes that we are stratifying (conditioning) on it.
- The dashed line denotes the induced conditional association.

How can two variables be associated in a sub-population? (2)

Some intuition



- Suppose there is a selective school that accepts pupils who are either good at sport, or good academically, or both.
- Suppose too that sporting ability and academic ability are independent in the population.
- Within this school, there will be a (negative) association between sporting and academic ability.
- Why? Suppose you choose a pupil at random and find her to be useless at sport. Then she must be good academically.

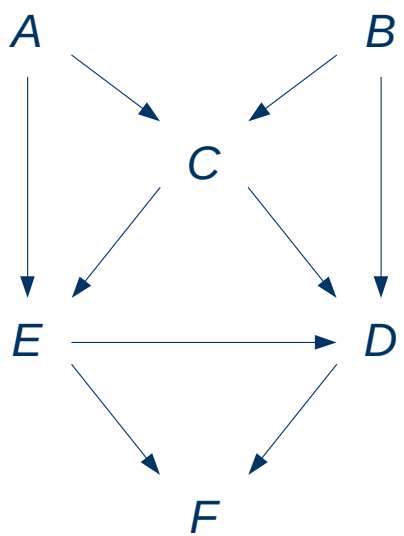
Summary so far

- X and Y will be associated in the population if:
  - X causes Y,
  - Y causes X, or
  - there is a Z that is a cause of both X and Y.
- X and Y will be associated in sub-populations defined by Z if Z is an effect of both X and Y.
- These are the building blocks of causal diagrams.

Outline

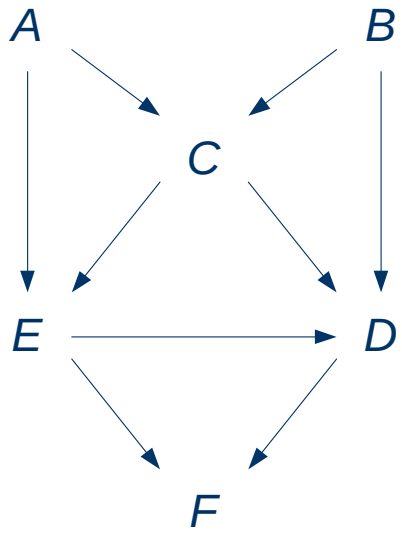
Conditional exchangeability and causal diagrams

- Conditional exchangeability is the key criterion that allows us to make causal statements using observational data
- Causal diagrams can help us identify the variables such that, conditional on them, exchangeability holds (given our assumptions!)
- The most commonly used causal diagrams are Directed Acyclic Graphs (DAGs):



- This DAG is a graphical representation of our causal assumptions
- It is made of *nodes* (i.e. variables) and *arrows*
- The arrows indicate the direction of causality.

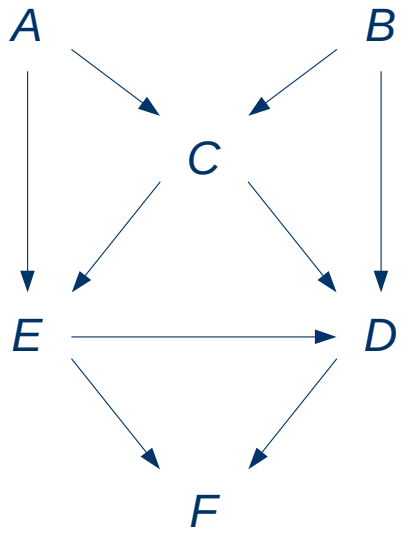
Causal Directed Acyclic Graph (DAG)



DAG stands for:

- D: directed since each edge is a single-headed arrow
- A: acyclic: it contains no cycles (no variable causes itself)
- G: graph

Causal Directed Acyclic Graph (DAG)



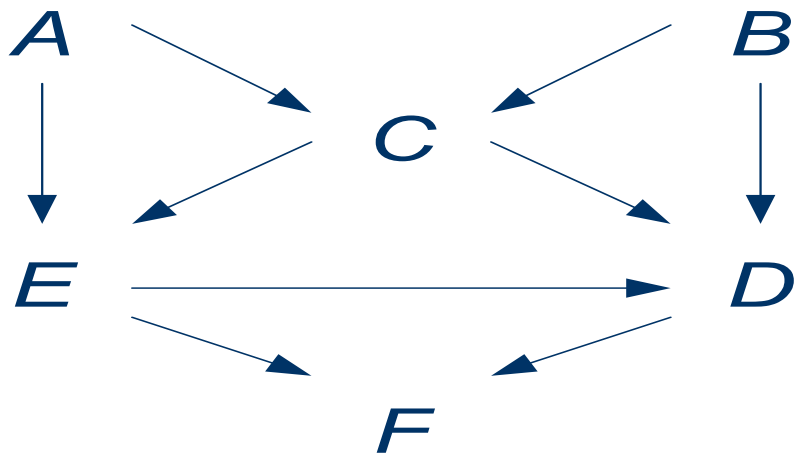
DAG stands for:

- D: directed since each edge is a single-headed arrow
- A: acyclic: it contains no cycles (no variable causes itself)
- G: graph

- Any common cause (even if unobserved) of two or more nodes in the DAG must itself be included in the DAG.
- Assumptions are encoded in the absence of arrows.

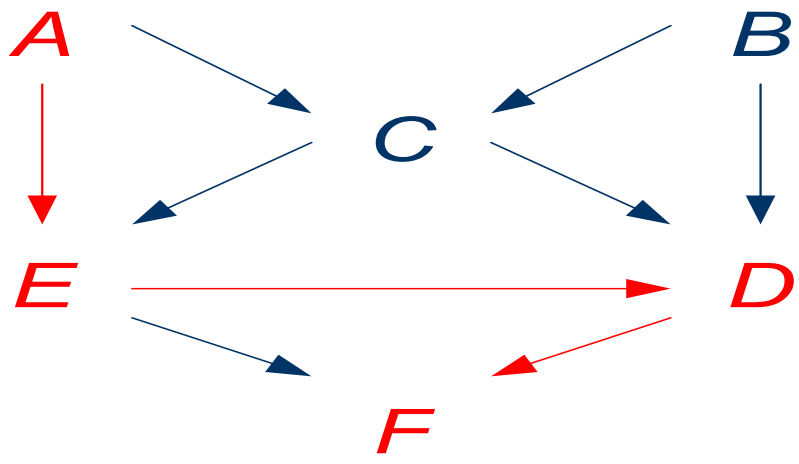
Terminology: nodes

Nodes in a path take various names. In particular:



- parent and child (as in **A** and **E**),
- ancestor and descendant (as in **A** and **F**)

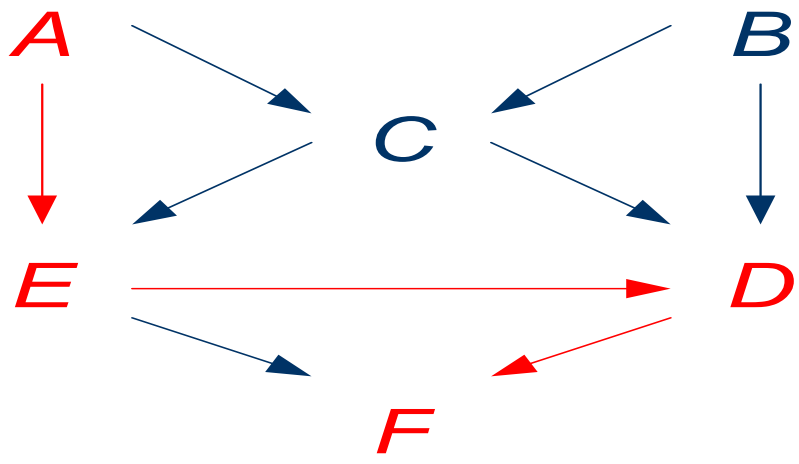
Terminology: paths



A path is a sequence of arrows pointing in any directions.

- directed path: path between any two variables where all arrows are single-headed and point 'forwards' (e.g. $\mathbf{A} \rightarrow \mathbf{E} \rightarrow \mathbf{D} \rightarrow \mathbf{F}$)

Terminology: paths

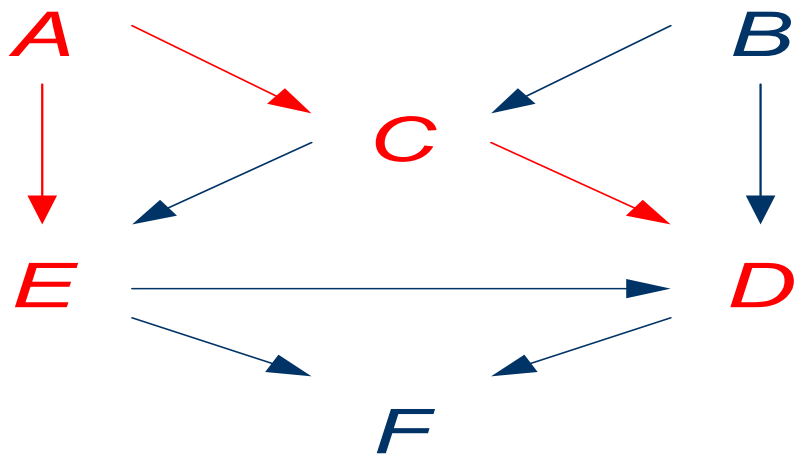


A path is a sequence of arrows pointing in any directions.

- directed path: path between any two variables where all arrows are single-headed and point 'forwards' (e.g. **A** $\rightarrow$ **E** $\rightarrow$ **D** $\rightarrow$ **F**)

It is a CAUSAL path because changing its start value changes the nodes that follow.

Terminology: paths



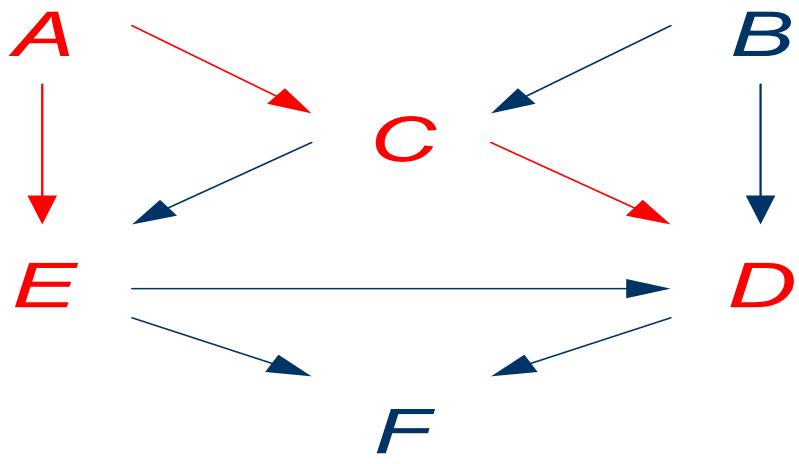
A path is a sequence of arrows pointing in any directions.

- directed path: path between any two variables where all arrows are single-headed and point 'forwards' (e.g. **A** $\rightarrow$ **E** $\rightarrow$ **D** $\rightarrow$ **F**)

It is a CAUSAL path because changing its start value changes the nodes that follow.

- back-door path: path between any two variables where the path starts with an arrow that points to the first (e.g. **E** $\leftarrow$ **A** $\rightarrow$ **C** $\rightarrow$ **D**)

Terminology: paths



A path is a sequence of arrows pointing in any directions.

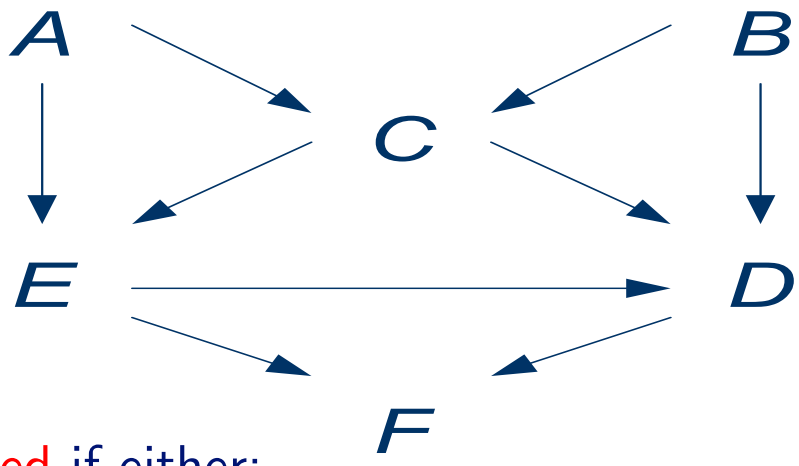
- directed path: path between any two variables where all arrows are single-headed and point 'forwards' (e.g. **A** $\rightarrow$ **E** $\rightarrow$ **D** $\rightarrow$ **F**)

It is a CAUSAL path because changing its start value changes the nodes that follow.

- back-door path: path between any two variables where the path starts with an arrow that points to the first (e.g. **E** $\leftarrow$ **A** $\rightarrow$ **C** $\rightarrow$ **D**)

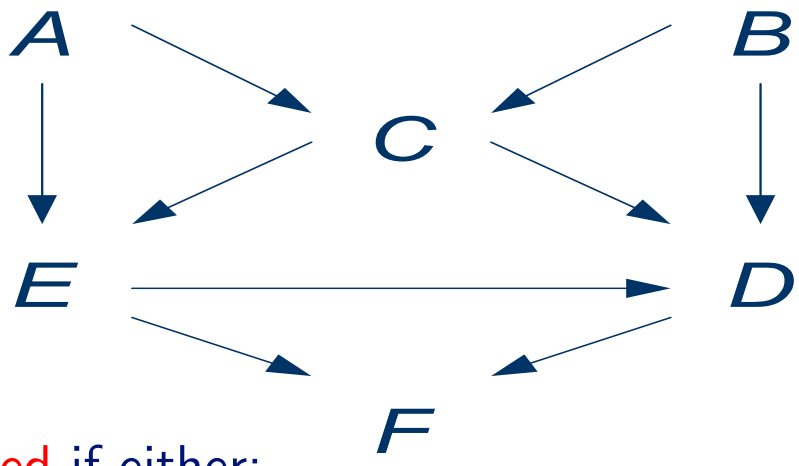
This is an ASSOCIATIONAL path, NOT a causal path.

Terminology: open and closed paths



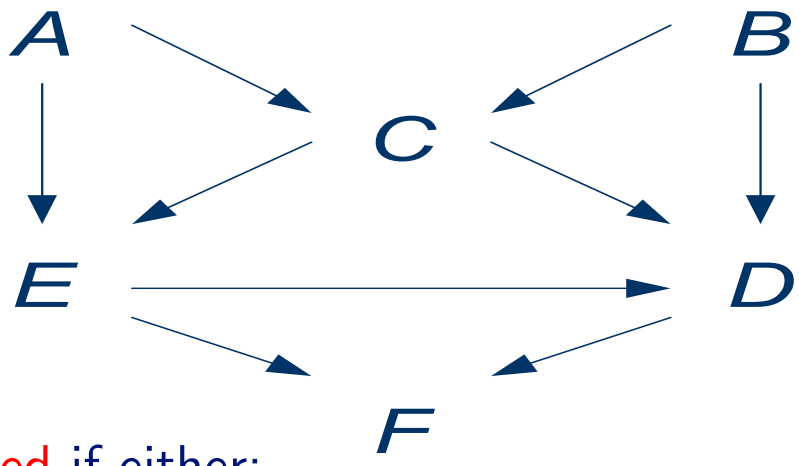
- A path is closed if either:
  a) there is at least one node in the path on which arrows converge, i.e. a collider, and we DO NOT condition on the collider nor any of its descendants, e.g. $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$

Terminology: open and closed paths



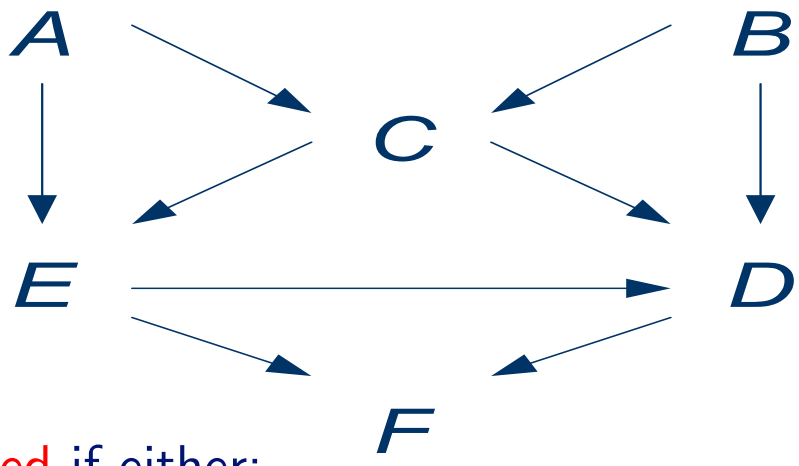
- A path is closed if either:
  a) there is at least one node in the path on which arrows converge, i.e. a collider, and we DO NOT condition on the collider nor any of its descendants, e.g. $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$
  b) we condition on at least one node in the path which is NOT a collider, e.g. we condition on A in the path $E \leftarrow A \rightarrow C \rightarrow D$.

Terminology: open and closed paths



- A path is closed if either:
  a) there is at least one node in the path on which arrows converge, i.e. a collider, and we DO NOT condition on the collider nor any of its descendants, e.g. $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$
  b) we condition on at least one node in the path which is NOT a collider, e.g. we condition on A in the path $E \leftarrow A \rightarrow C \rightarrow D$.
- A path is open if it is not closed.

Terminology: open and closed paths

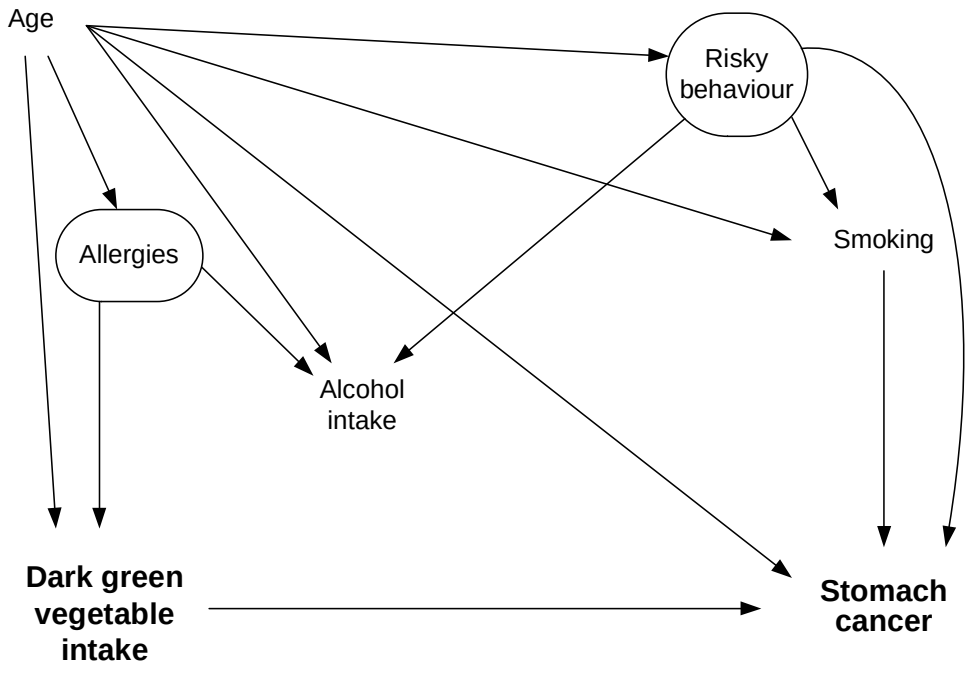


- A path is closed if either:
  a) there is at least one node in the path on which arrows converge, i.e. a collider, and we DO NOT condition on the collider nor any of its descendants, e.g. $E \leftarrow A \rightarrow C \leftarrow B \rightarrow D$
  b) we condition on at least one node in the path which is NOT a collider, e.g. we condition on A in the path $E \leftarrow A \rightarrow C \rightarrow D$.
- A path is open if it is not closed.
- Open paths transmit associations; closed paths do not.

Summary so far

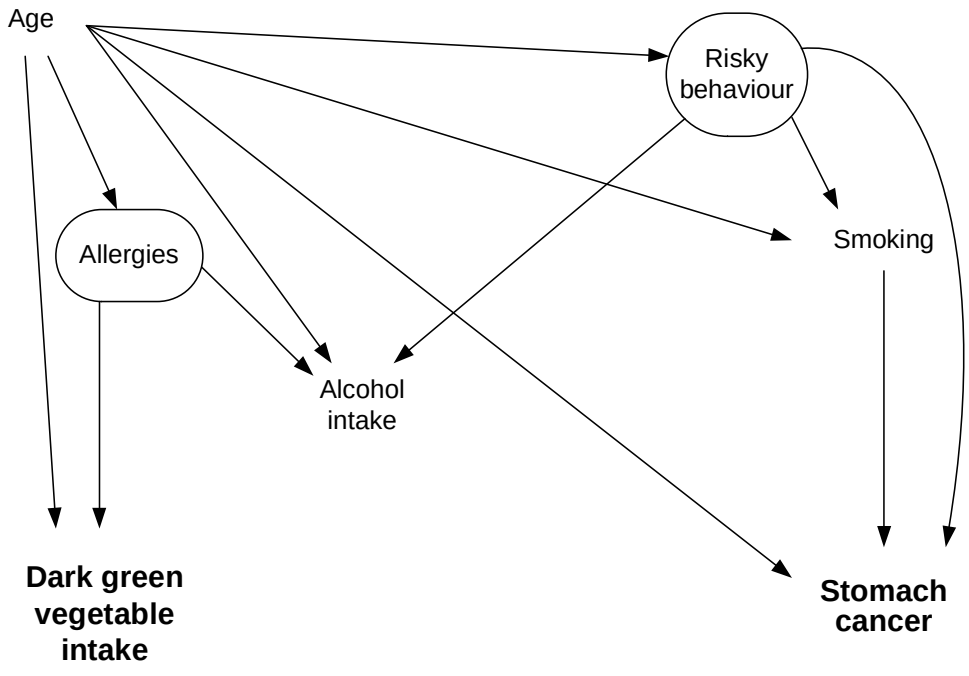
- We are after the causal effect of **E** on **D**.
- We now know how to draw a DAG that represents our causal assumptions
- and to classify all paths from **E** to **D**.
- We know that if there are non-causal paths that are open, exposed and unexposed individuals are not exchangeable.
- Next we must identify the nodes (i.e. variables) needed to block all open paths that are non-causal.

The example



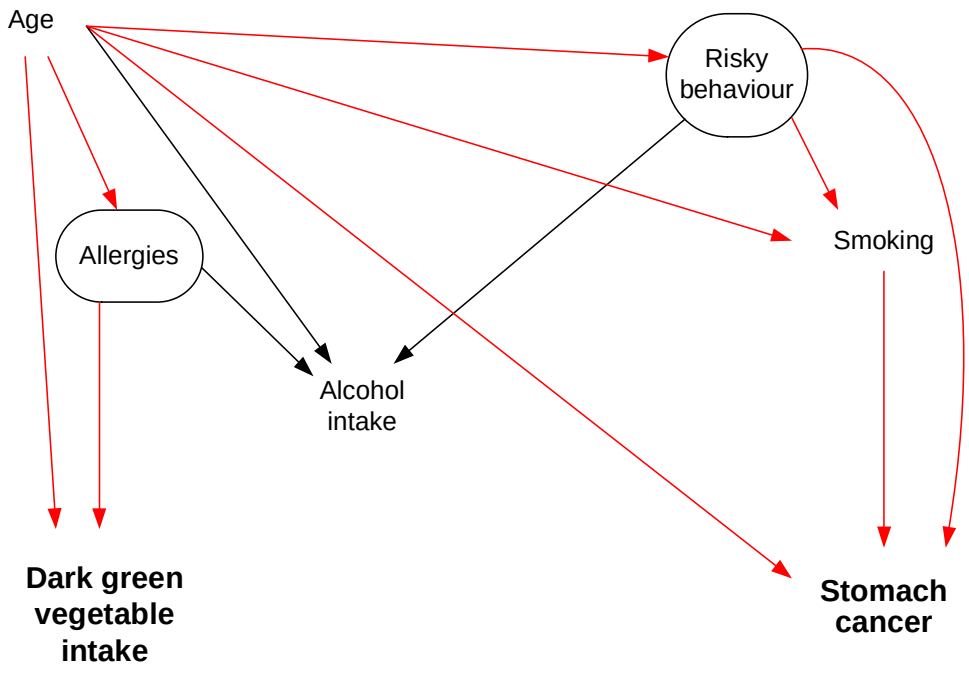
Here is (one of many possible) causal diagrams for our motivating example.

The example



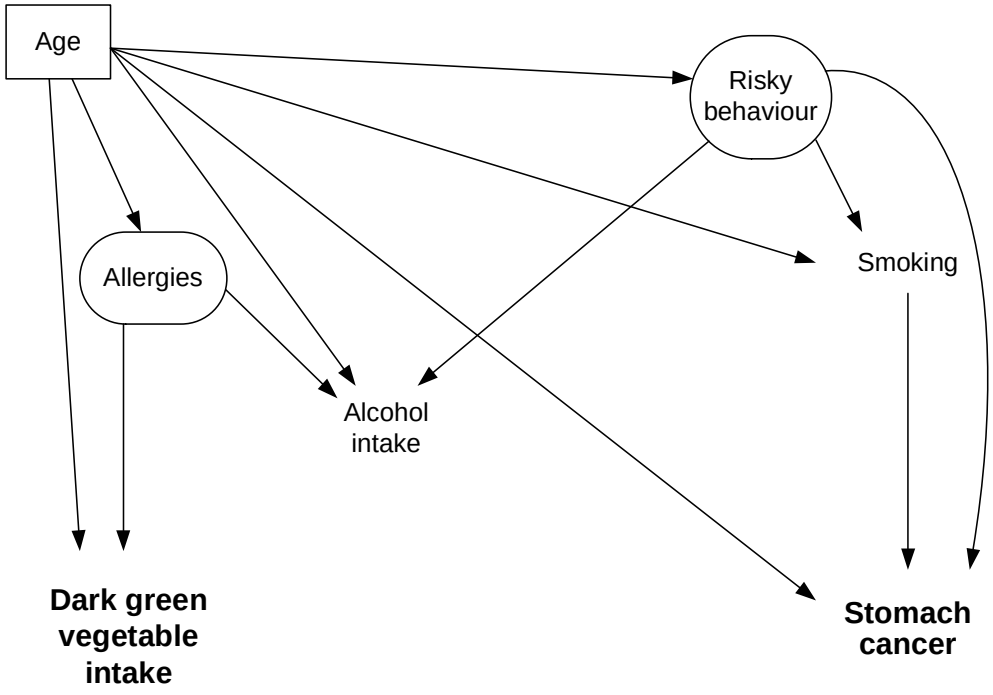
To facilitate our reading of the DAG, remove the arrow from exposure to outcome.

The example



- Many back-door paths exist.
- They all pass through age.
- Those shown in red are all open.

The example

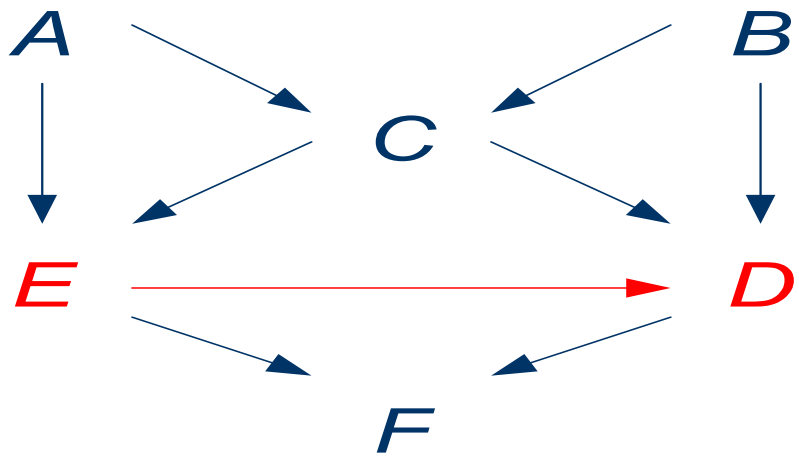


- Conditioning on age would block all of them.

Outline

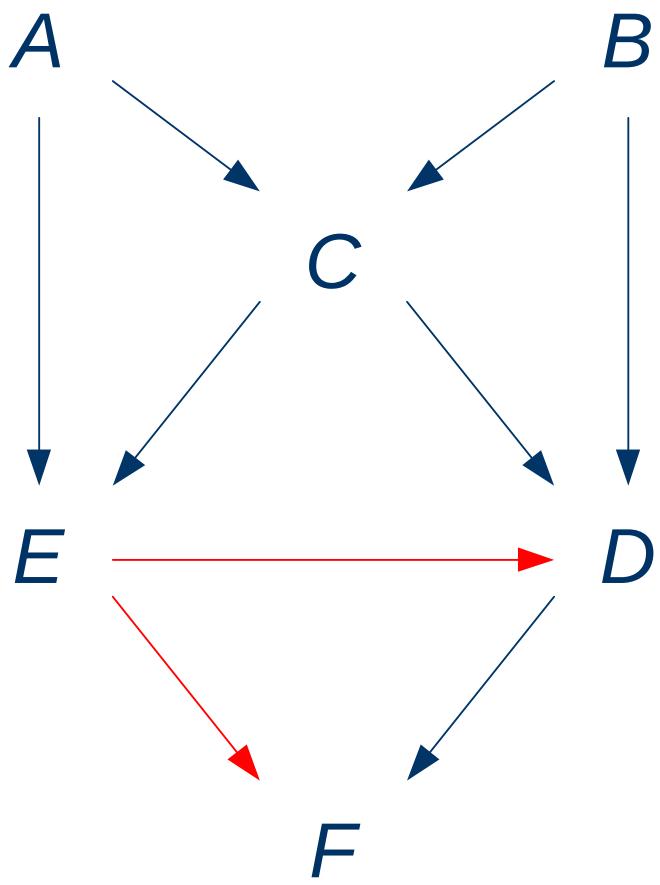
Back-door criterion

What we have done now is a simple application of the back-door criterion.
More formally, and using a general notation for an exposure E and outcome D,



The back-door criterion consists of two steps.

Back-door criterion (Step 1)

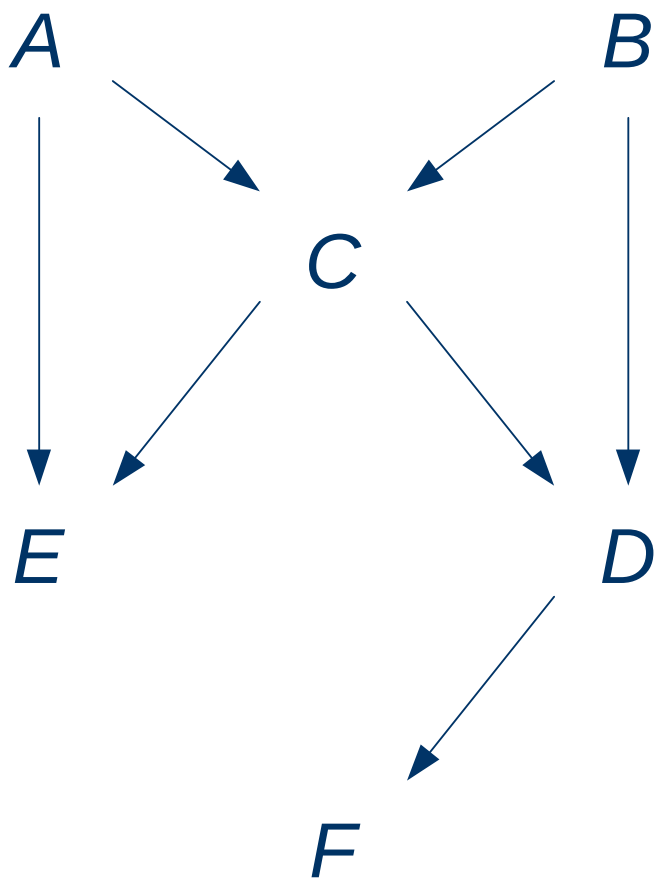


Step 1

- First we remove all arrows emanating from the exposure.

We do this because paths out of E either causally lead to D or, if not, they are already blocked by a collider.

Back-door criterion (Step 2)



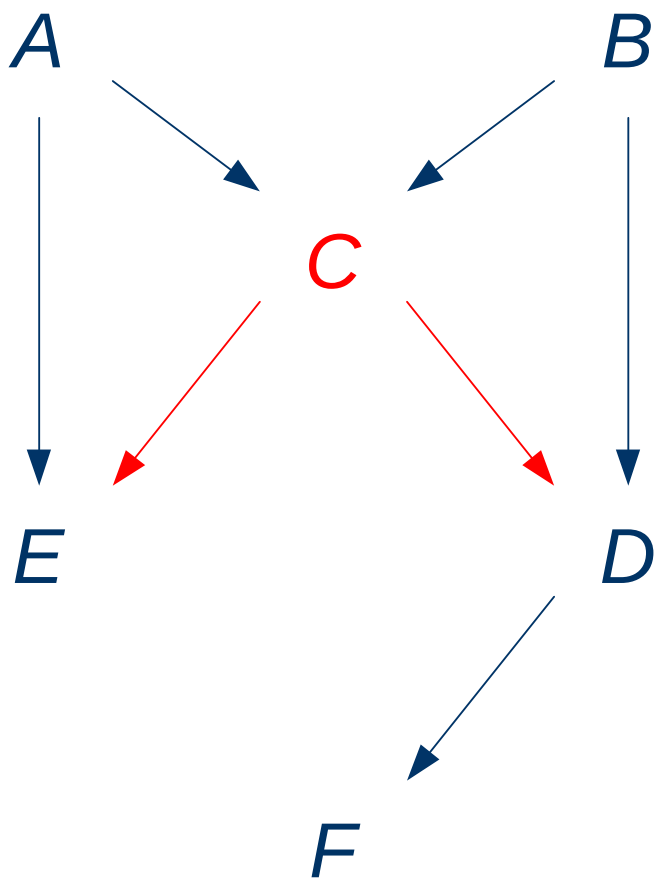
Step 2

- Then we look for any open paths from the exposure to the outcome.

Recall: an open path does not contain a collider.

Back-door criterion

Check 1

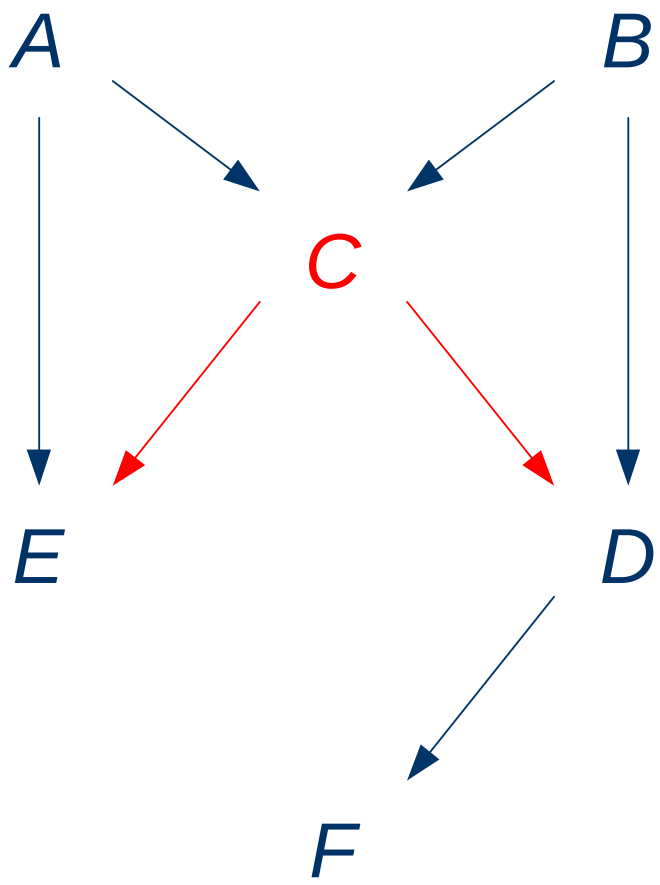


Step 2

- Is this an open path?

Back-door criterion

Check 1

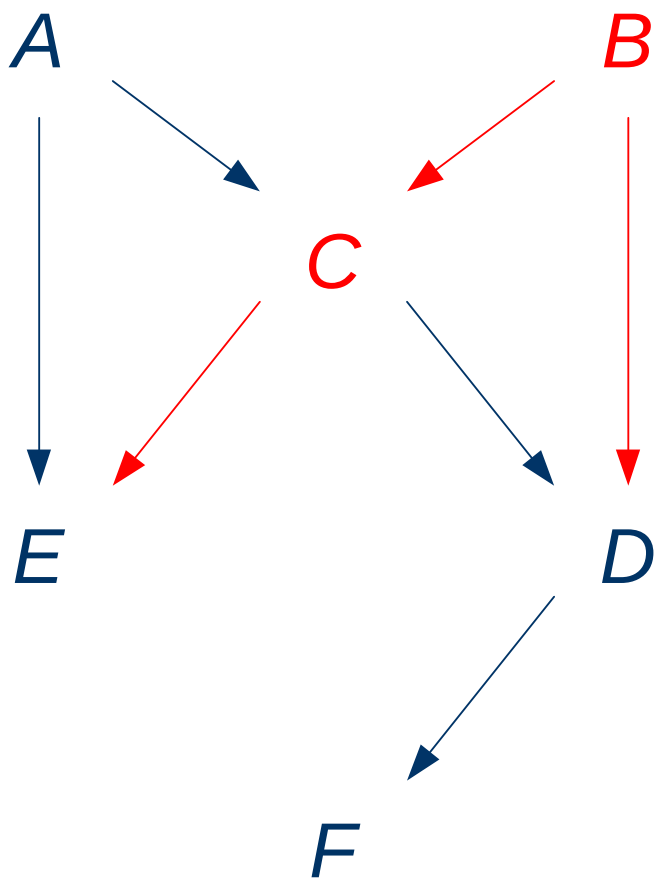


Step 2

- Is this an open path?
- Yes.

Back-door criterion

Check 2

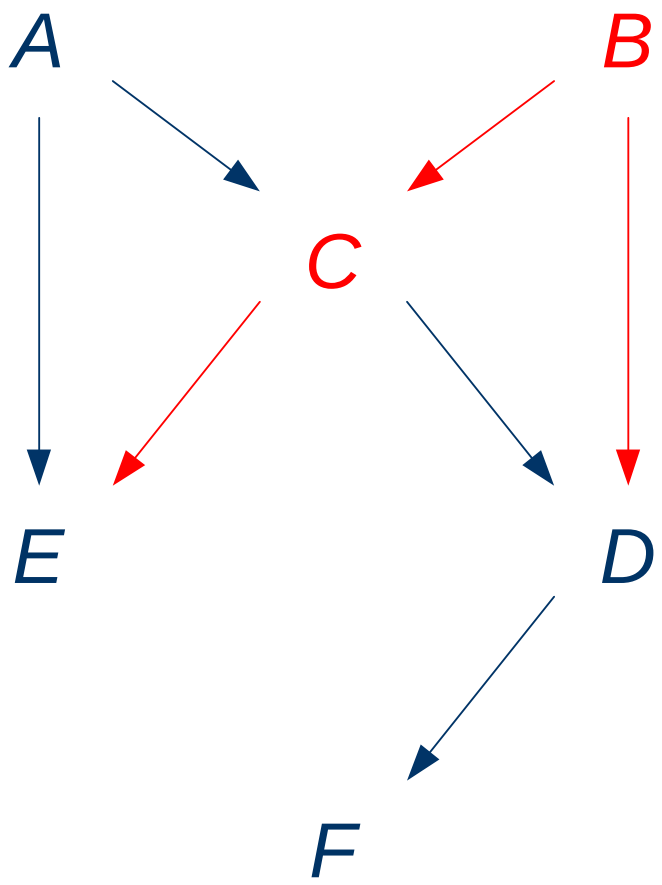


Step 2

- Is this an open path?

Back-door criterion

Check 2

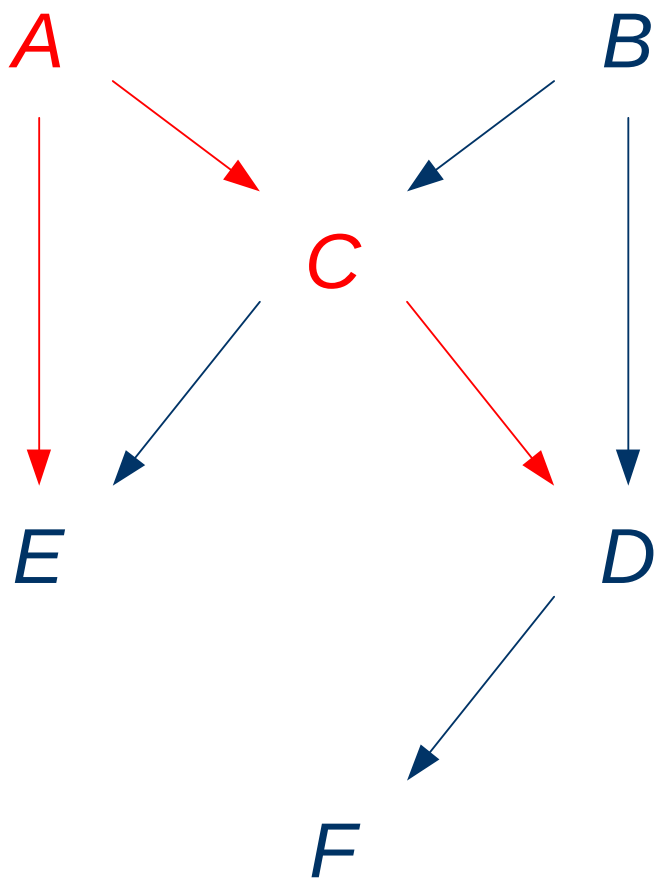


Step 2

- Is this an open path?
- Yes.

Back-door criterion

Check 3

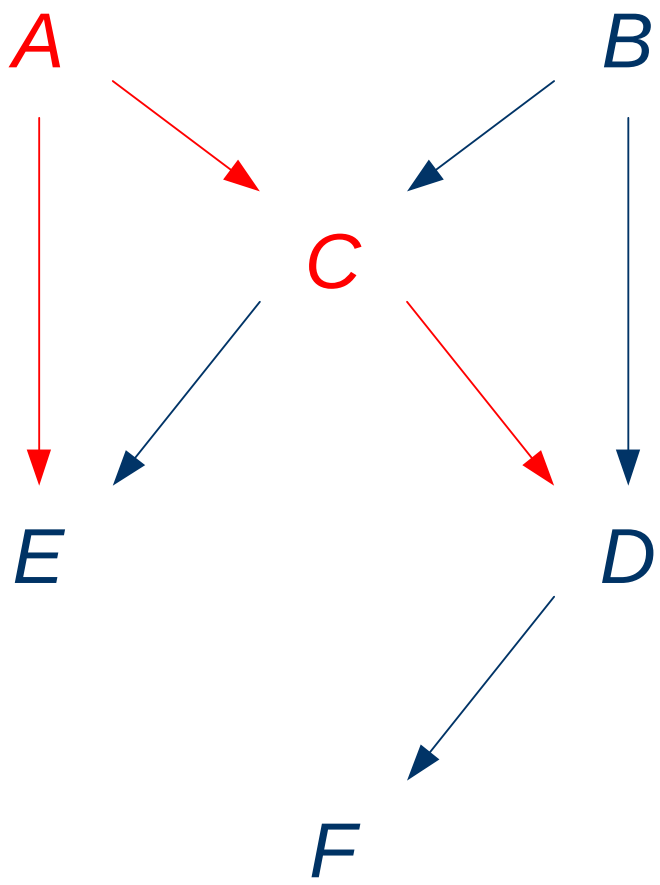


Step 2

- Is this an open path?

Back-door criterion

Check 3

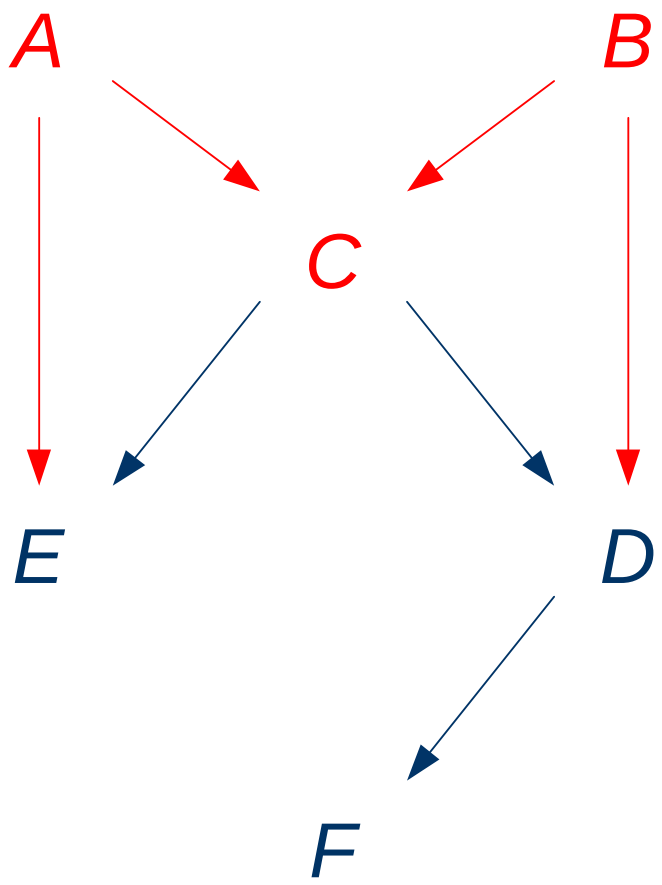


Step 2

- Is this an open path?
- Yes.

Back-door criterion

Check 4

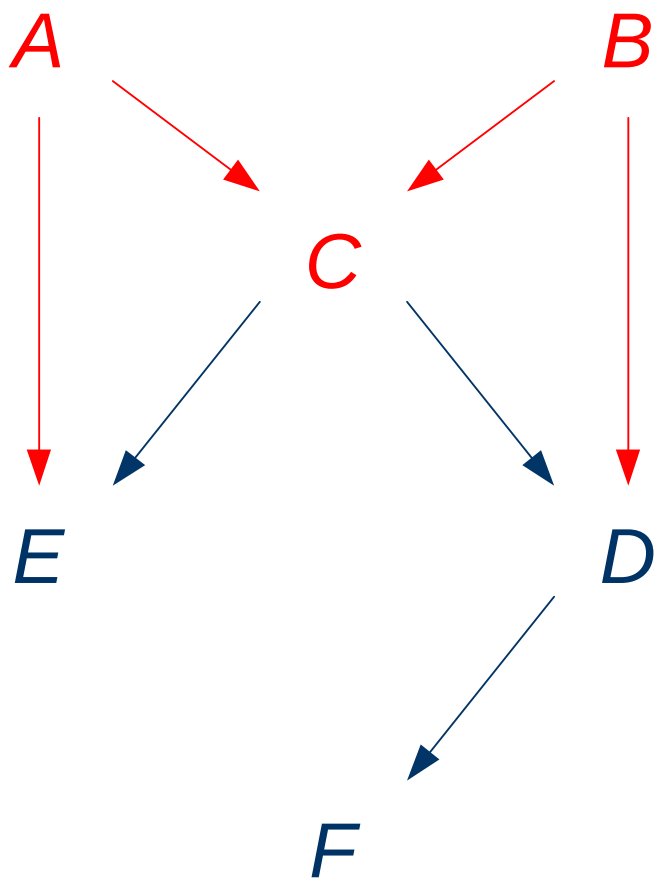


Step 2

- Is this an open path?

Back-door criterion

Check 4

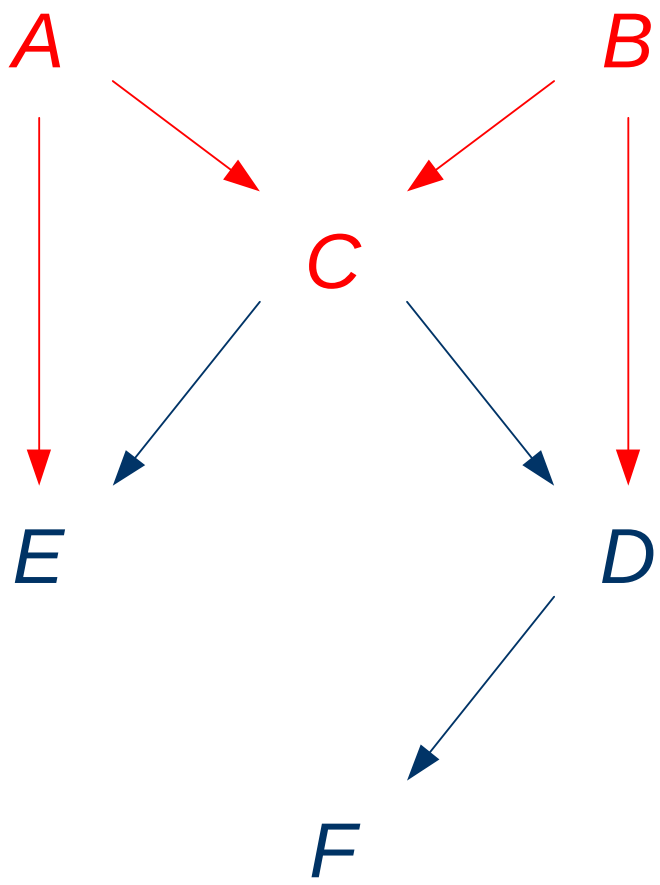


Step 2

- Is this an open path?
- No!

Back-door criterion

Check 4



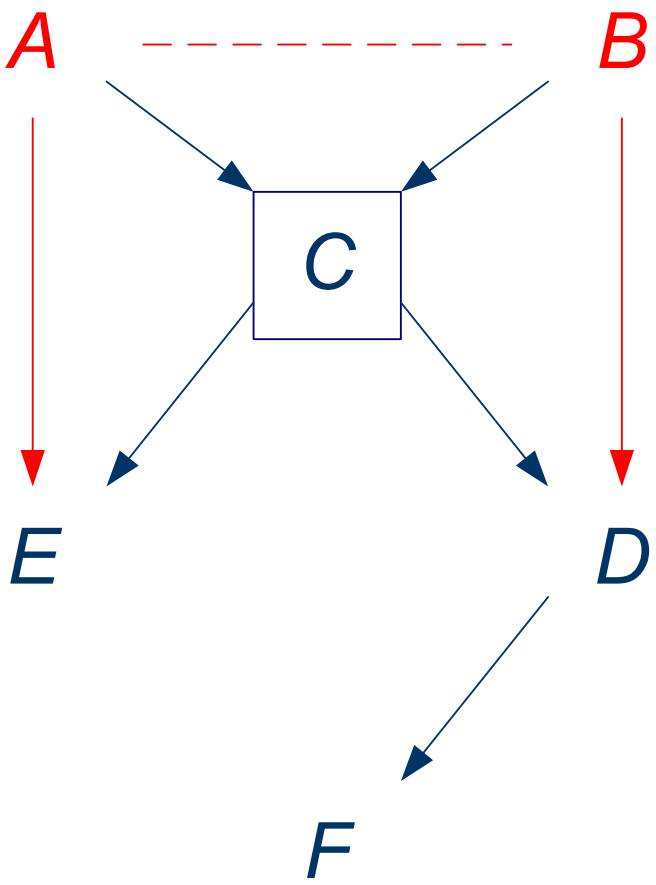
Step 2

- Is this an open path?
- No!

What happens if we condition on C?

Back-door criterion

Check 4



Step 2

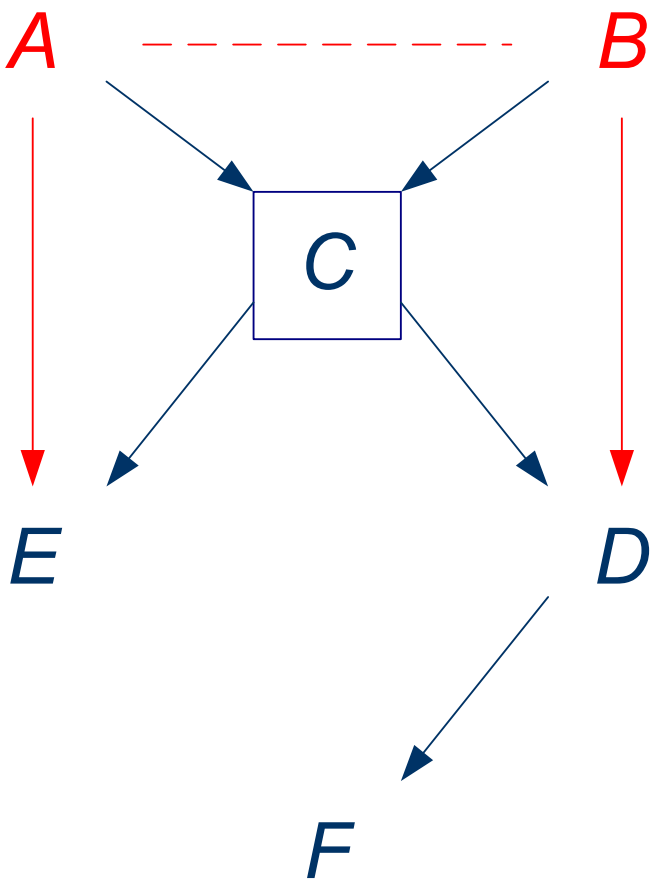
- Is this an open path?
- No!

What happens if we condition on C?

- We create a new open path!

Back-door criterion

Check 4



Step 2

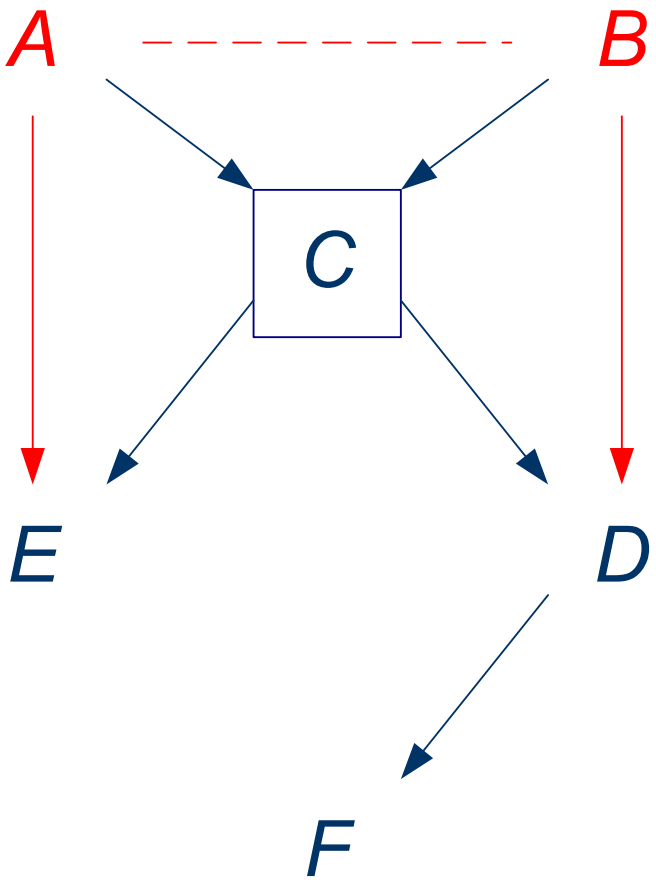
- Is this an open path?
- No!

What happens if we condition on C?

- We create a new open path!
- What can we do?

Back-door criterion

Check 4



Step 2

- Is this an open path?
- No!

What happens if we condition on C?

- We create a new open path!
- What can we do?
- We can condition on A or B to close the path opened by conditioning on C.

The back-door algorithm

The back-door criterion allows us to:

1. identify all the open back-door paths ← just seen

The back-door algorithm

The back-door criterion allows us to:

1. identify all the open back-door paths ← just seen
2. determine whether a set of covariates is sufficient to block them.

The back-door algorithm

The back-door criterion allows us to:

1. identify all the open back-door paths ← just seen
2. determine whether a set of covariates is sufficient to block them.

More formally, once we have selected a set of covariates $\mathcal{S}$:

The back-door algorithm

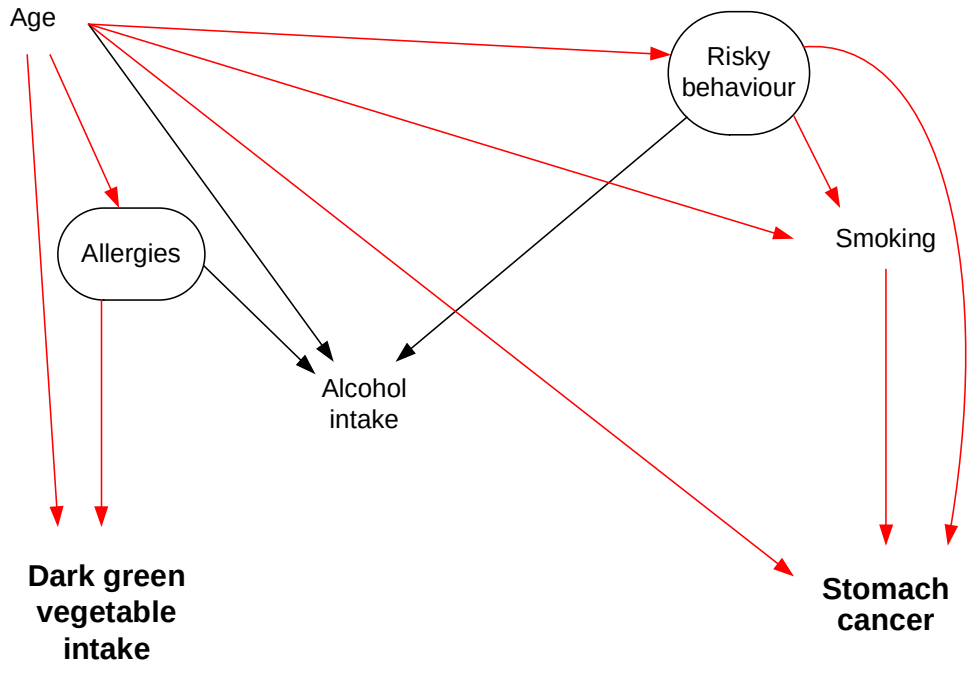
(i) we make sure that $\mathcal{S}$ does not contain any descendants of the exposure

(ii) we remove all arrows emanating from the exposure

(iii) we join with a dotted line any two variables that share a child which is either itself in $\mathcal{S}$ or has a descendant in $\mathcal{S}$

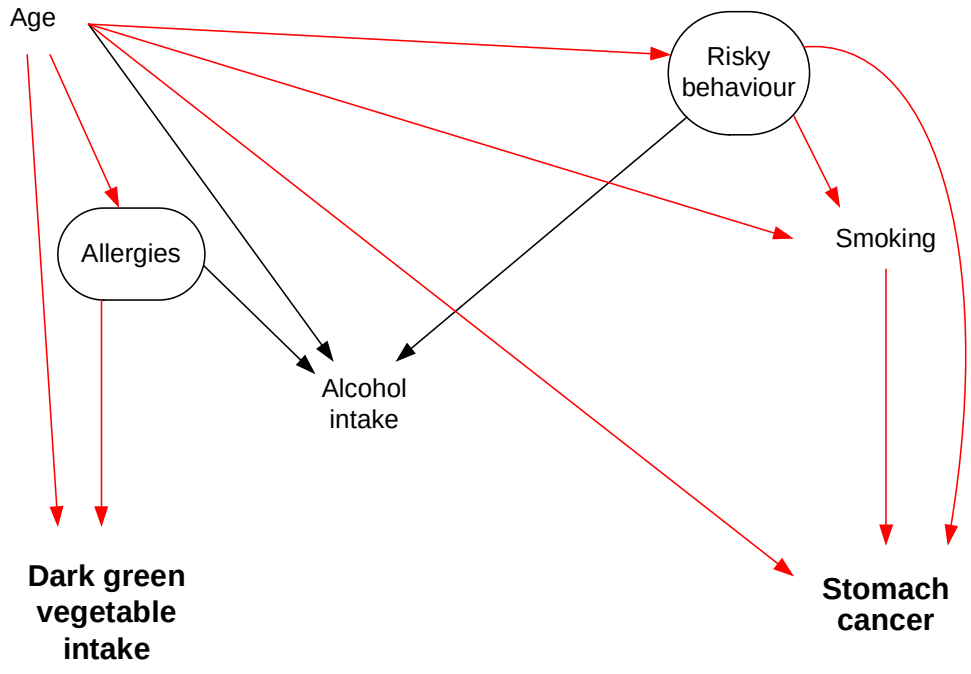
(iv) we ask: "Is there an open path from E to D that does not pass through a member of $\mathcal{S}$"?
If NOT, then $\mathcal{S}$ is sufficient to control for the confounding.

Revisiting the example



After removing all arrows emanating from the exposure, we see that there are many open back-door paths (in red).

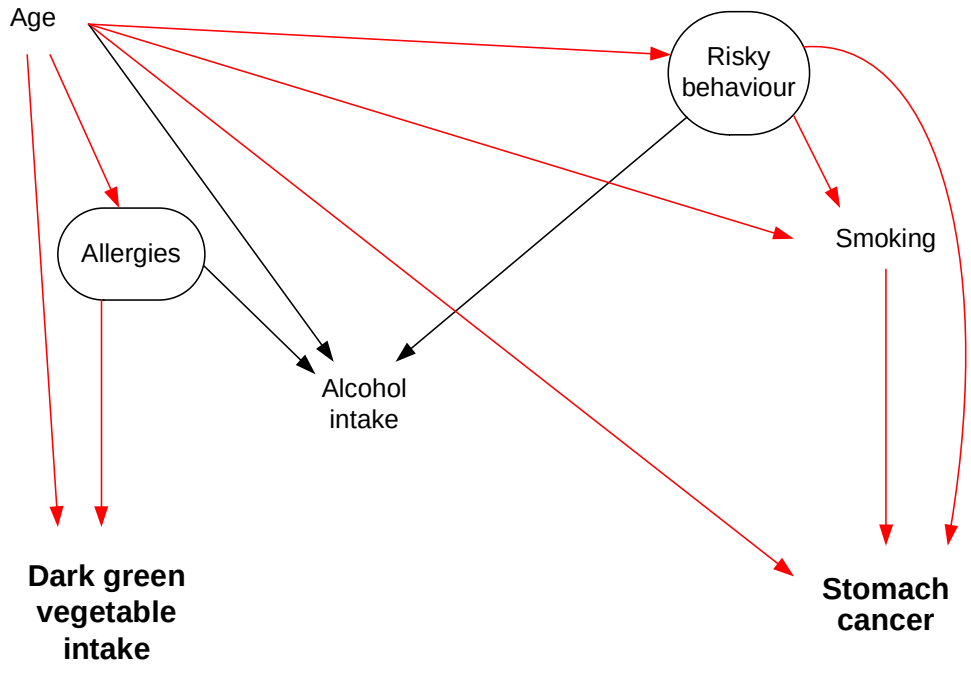
Revisiting the example



After removing all arrows emanating from the exposure, we see that there are many open back-door paths (in red).

- Choose age as the candidate member of the set $\mathcal{S}$.

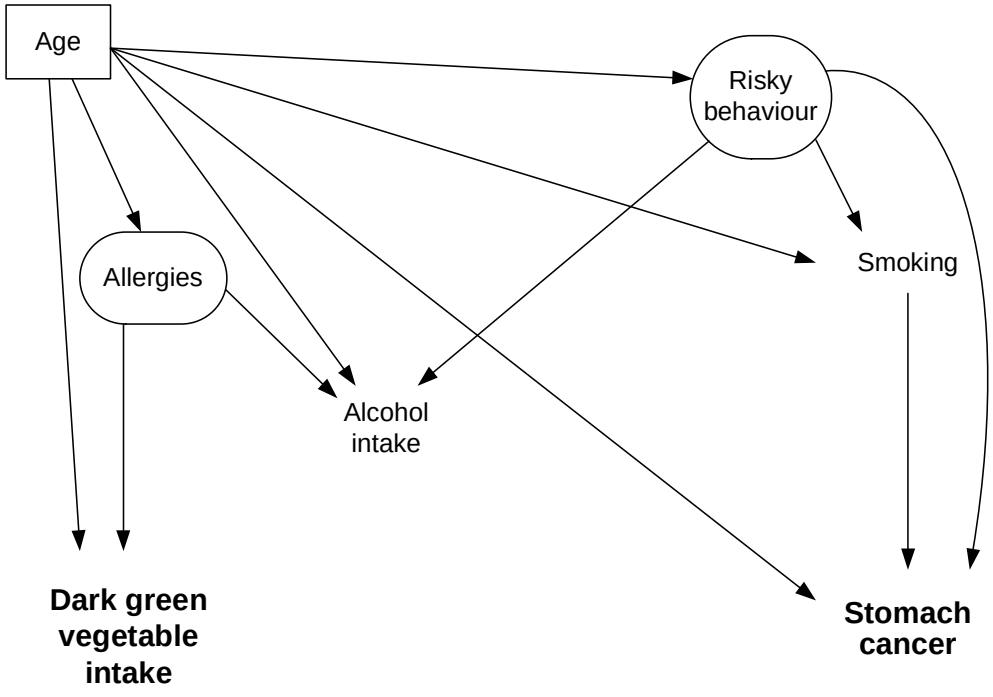
Revisiting the example



After removing all arrows emanating from the exposure, we see that there are many open back-door paths (in red).

- Choose age as the candidate member of the set $\mathcal{S}$.
- Is $\mathcal{S}$ sufficient?

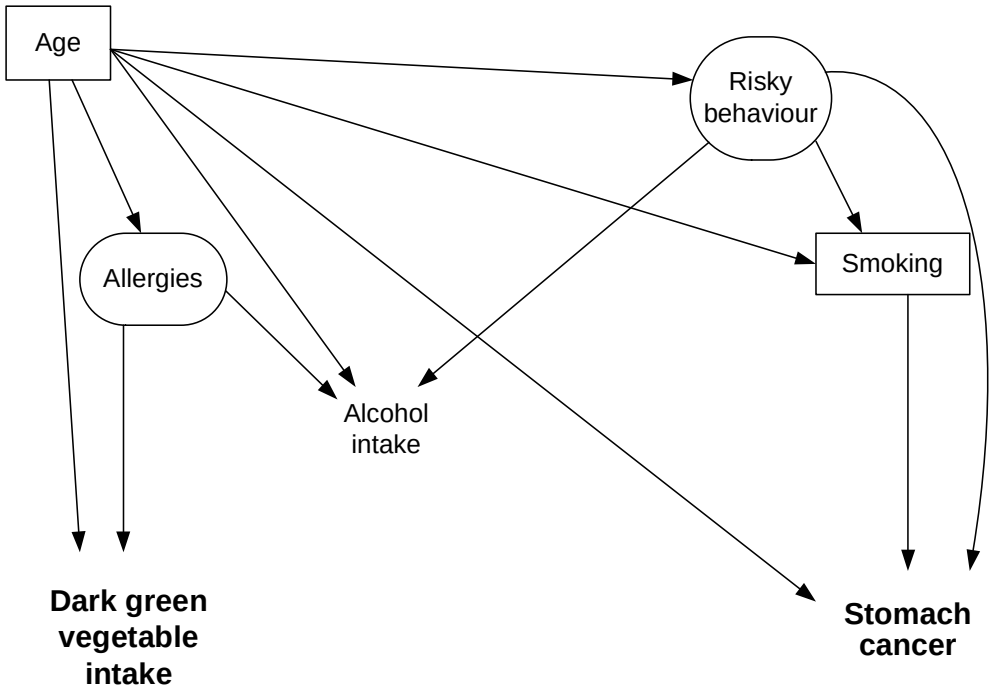
Revisiting the example



After removing all arrows emanating from the exposure, we see that there are many open back-door paths (in red).

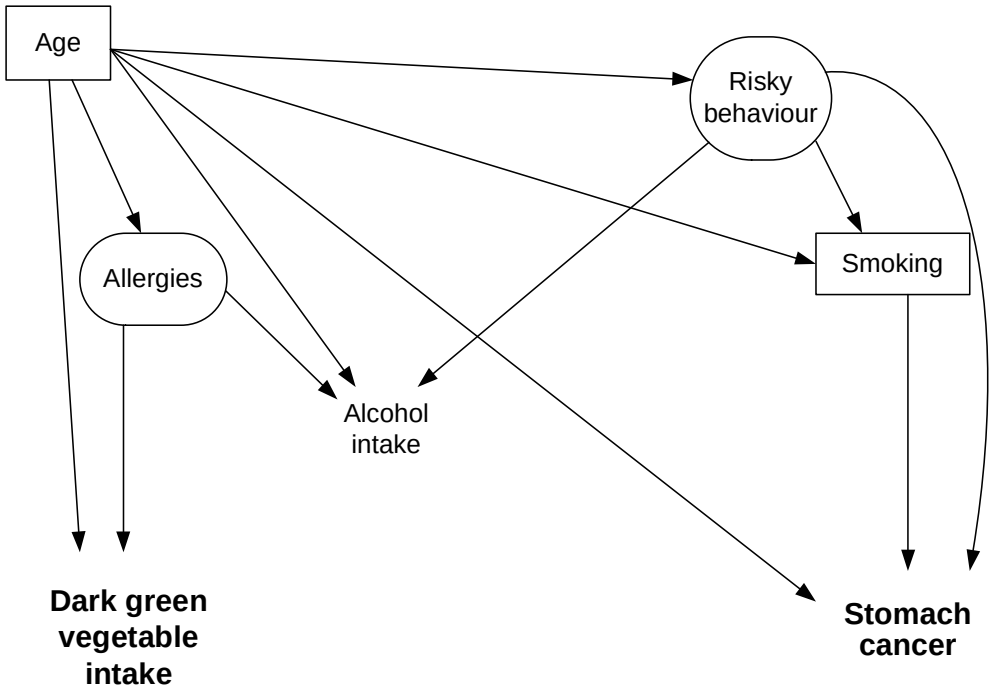
- Choose age as the candidate member of the set $\mathcal{S}$.
- Is $\mathcal{S}$ sufficient? Yes because it blocks all back-door paths.

Revisiting the example



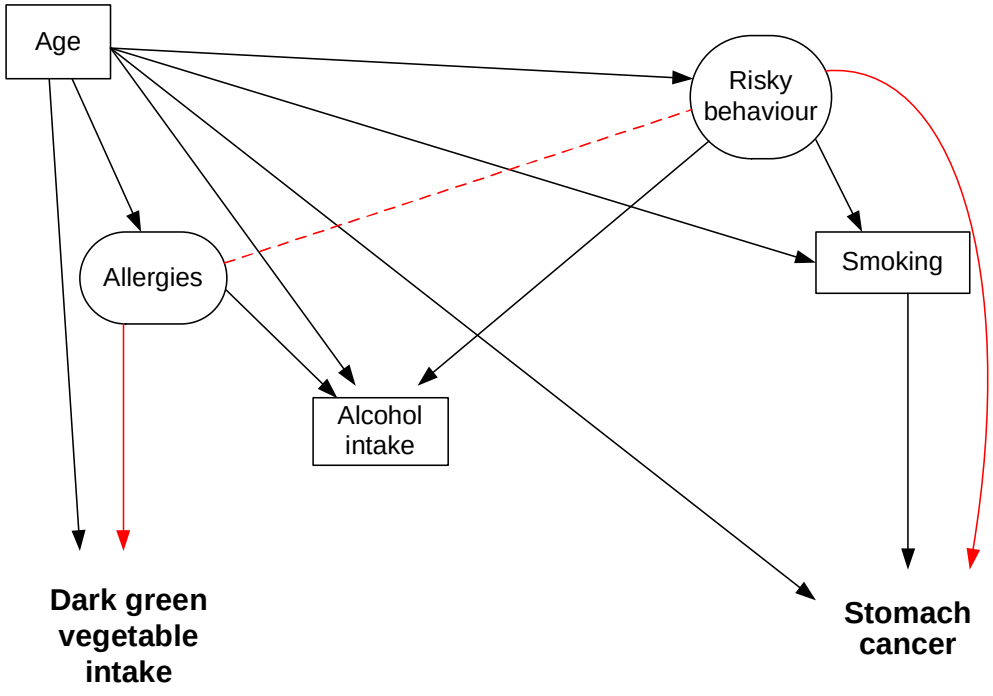
- Now add smoking to the set $\mathcal{S}$ that already includes age.
- Is $\mathcal{S}$ sufficient?

Revisiting the example



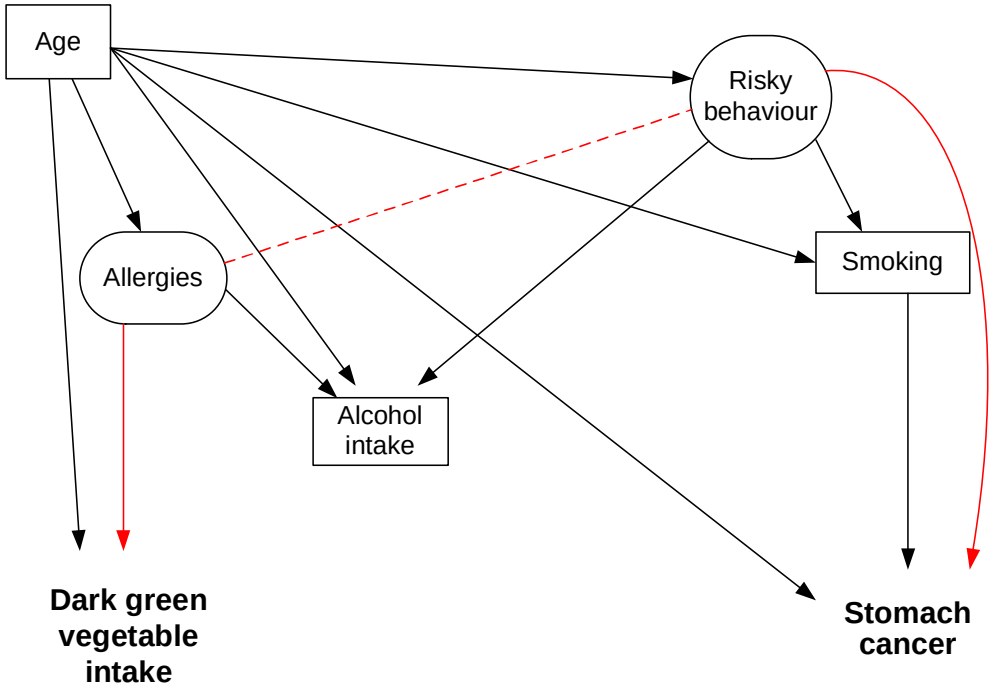
- Now add smoking to the set $\mathcal{S}$ that already includes age.
- Is $\mathcal{S}$ sufficient? Yes, there are no open back-door paths.

Revisiting the example



- Finally add alcohol to the set $\mathcal{S}$ that already includes age and smoking.
- Is $\mathcal{S}$ sufficient?

Revisiting the example



- Finally add alcohol to the set $\mathcal{S}$ that already includes age and smoking.
- Is $\mathcal{S}$ sufficient? No!!!! We have now opened a new back-door path. Unfortunately we cannot block it because we cannot measure either 'Allergies' or 'Risky behaviour'.

DAGitty: a useful tool

- Applying the back-door criterion can become very tedious by hand, especially in situations with many variables.
- A free web-based computer program exists to implement the algorithm: `www.dagitty.net`.
- Let's attempt a live demonstration...

The intuition behind the back-door criterion

- Hopefully, the back-door criterion algorithm makes sense in the light of our basic building blocks:
  1. An association is transmitted along an open path unless we condition (stratify) on a variable along that path.
  2. An association is NOT transmitted along a blocked path UNLESS we condition on a variable along it, or a descendant of a variable along it.
- The back-door criterion asks "after conditioning on $\mathcal{S}$, and in the absence of a causal effect of E on D, would we still see an association between E and D?". If YES, then $\mathcal{S}$ is not sufficient, and there is still confounding even if we control for $\mathcal{S}$.

Outline

Analyses following the back-door criterion

- Suppose we have (carefully) drawn our causal diagram, applied the back-door criterion and identified a sufficient set of variables to control for confounding. What next?
- If the number of confounders is small and categorical/binary, we could stratify on them.
- We would then calculate our effect of interest (e.g. an odds ratio) in each stratum and then combine these in the usual way (Mantel–Haenszel), or report them separately if there are effect modifiers of interest, etc.
- Or, if there are too many confounders and/or some are continuous, we could specify a suitable regression model (linear/logistic/Poisson/Cox. . .):

$$\text{logit}\left\{Pr\left(\text{D}=1\,|\,E,\text{gender},\text{age}\right)\right\}=\alpha+\beta E+\gamma\text{gender}+\delta\text{age}$$

Is this correctly specified?

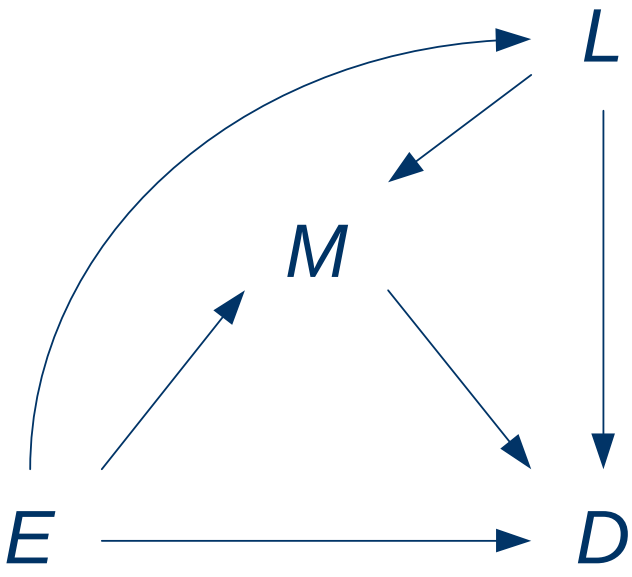
Alternative methods based on the propensity score

- There are other options.
- Instead of modelling the outcome given the exposure and confounders, we can model the outcome given the exposure (only), AND model the exposure given the confounders (and then stratify, match or re-weight using predictions from this model).
- Such methods are called propensity score methods.
- Either way, we use data on the sufficient set of confounders identified using the back-door criterion.

Approaches based on alternative causal assumptions

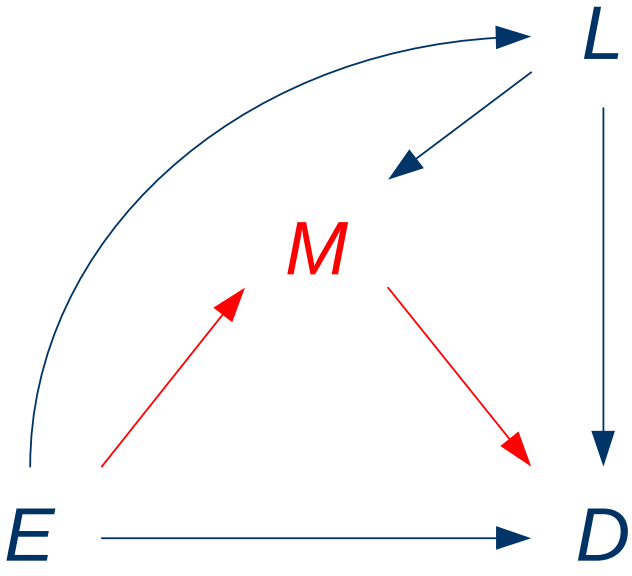
- Having drawn our causal diagram, we may find it thoroughly implausible that we have collected data on a sufficient set of confounders such that conditional exchangeability holds.
- In this situation, we can sometimes still make causal inferences, if we are prepared to make an alternative set of assumptions, namely that we have collected data on an instrumental variable for our exposure–outcome relationship.
- This assumption points towards an alternative set of statistical methods.

Pathway-specific effects (1)



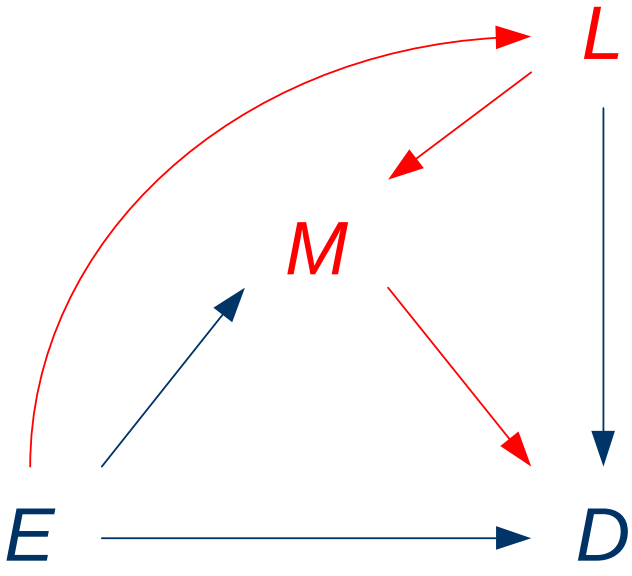
- Sometimes we are interested in more complex causal questions than simply "what is the (total) causal effect of (one exposure) E on (one outcome) D?"
- Take for example this somewhat complicated 'network'.
- E affects D 'directly' and 'indirectly', through M and L.

Pathway-specific effects (2)



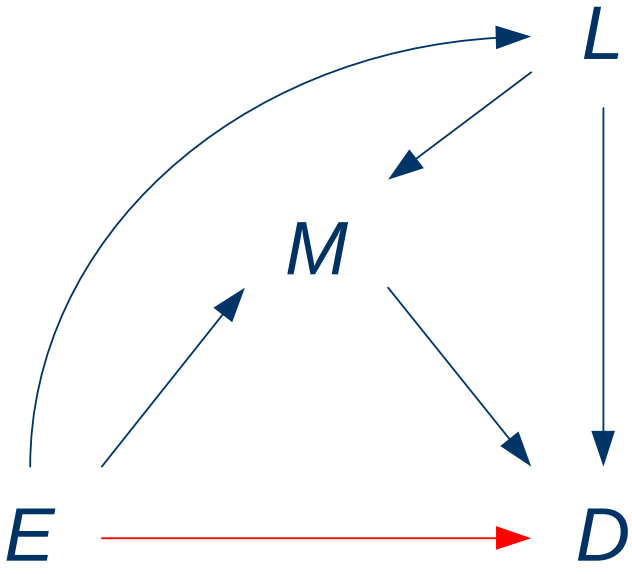
- For example, how much of the causal effect of E on D is mediated by M?

Pathway-specific effects (3)



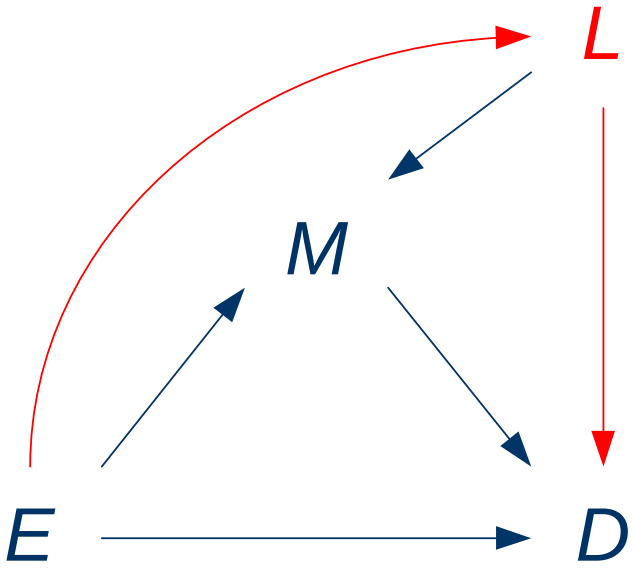
- For example, how much of the causal effect of E on D is mediated by M?

Pathway-specific effects (4)



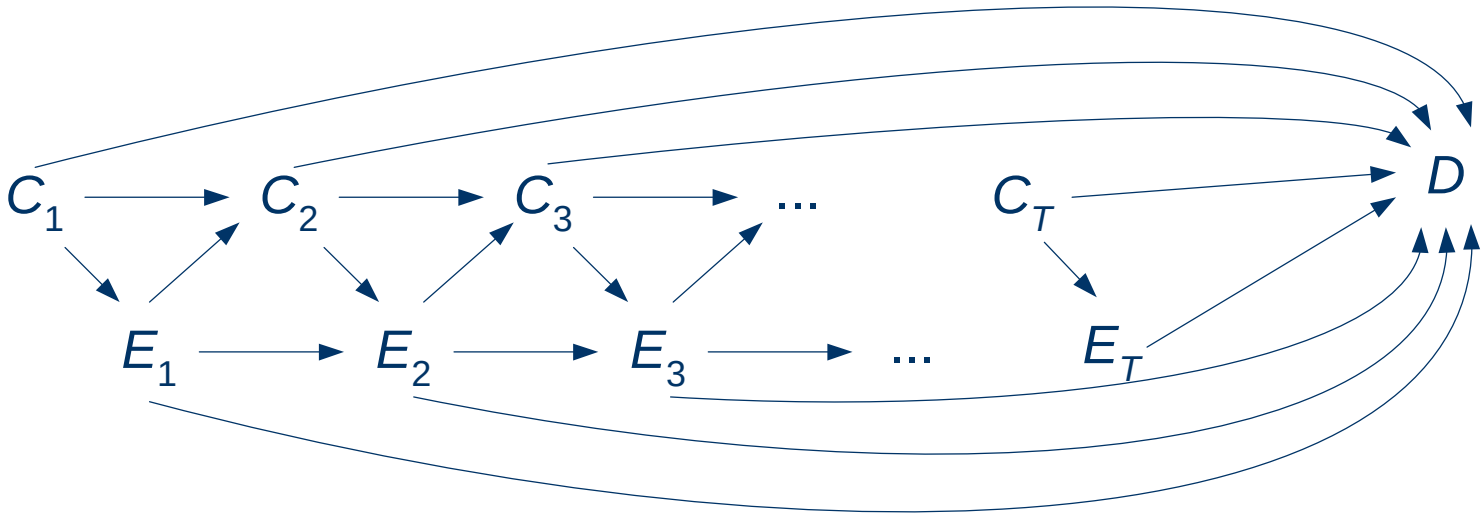
- And how much is not mediated by M?

Pathway-specific effects (5)



- And how much is not mediated by M?

Longitudinal exposures



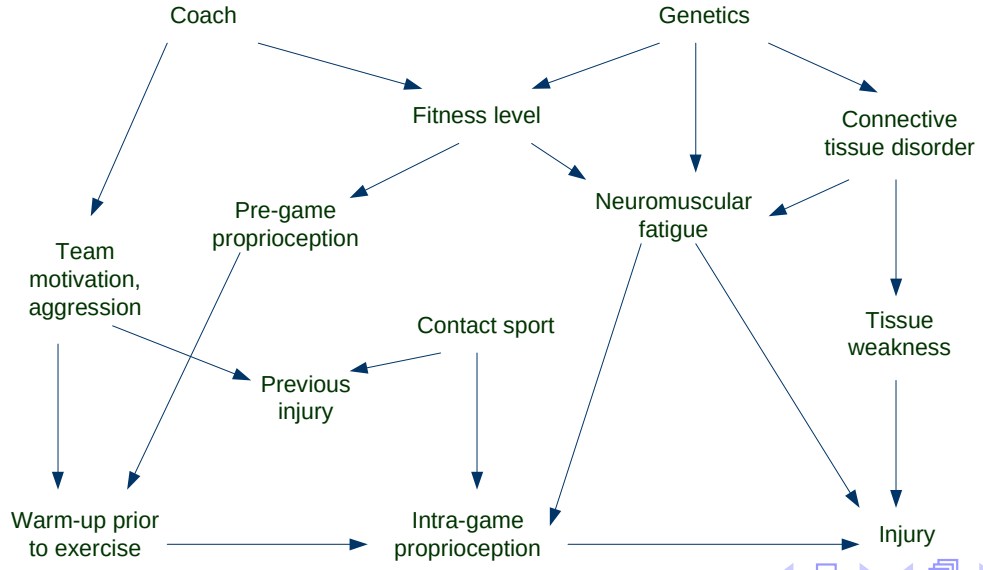
- Or, what is the joint causal effect of the time-varying exposures E_1, E_2, ... on D when there is causal feedback between E_1, E_2, ... and C_1, C_2, ...
- These are all questions that can (only) be addressed (given certain assumptions) using novel statistical methods.

Outline

Final remarks

Causal thinking:

- clarifies our research objectives
- reduces the ambiguities of traditional tools such as the epidemiological triangle
- shows us that conditioning can be harmful as well as helpful.
- allows us to deal with arbitrarily complex settings.
  The triangle would not take us very far in this example!



Summary

We aimed to:

1. Review traditional methods and highlight when they might be inadequate
2. Introduce causal language and demonstrate it is necessary if we wish to make causal statements
3. Define some essential tools (DAGs) and illustrate their use
4. Give a brief overview of the choice of existing and **new** statistical methods
5. Broaden the debate . . . !!!

Want to know more?

- Is there sufficient interest for future (more advanced) workshops?
- Visit the causal inference theme page on the LSHTM's Centre for Statistical Methodology website: `http://csm.lshtm.ac.uk/themes/causal-inference/`.
- Places are available for the 2011 short course *Causal Inference in Epidemiology: Recent Methodological Developments* to be held at LSHTM in the November reading week. For more info, see: `http://www.lshtm.ac.uk/prospectus/short/causal_inference.html`.

References (1)


Greenland, S., Pearl, J. and Robins, J. M. (1999)
Causal diagrams for epidemiologic research.
Epidemiology, 10:37–48.


Hernán, M. A., Hernández-Díaz, S., Werler, M. M. and Mitchell, A. A. (2002)
Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology.
American Journal of Epidemiology, 155:176–184.


Hernán, M. A. (2004)
A definition of causal effect for epidemiological research.
Journal of Epidemiology and Community Health, 58:265–271.


Maldonado, G. and Greenland, S. (2002)
Estimating causal effects.
International Journal of Epidemiology, 31:422–429.

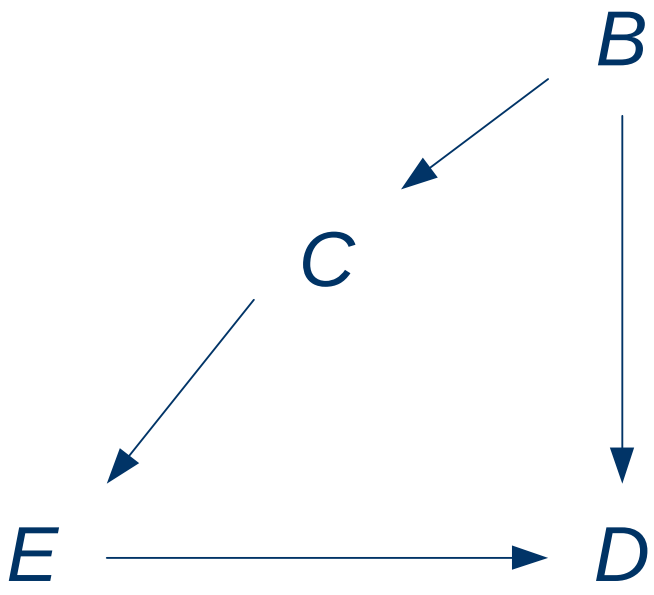
References (2)

Pearl, J. (2009)
Causality.
Cambridge University Press, 2nd ed.

Hernán, M. A. and Robins, J. M. (2011)
Causal Inference.
Chapman & Hall, not yet available, but draft chapters can be downloaded (for free) from `http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/`.

Hernán, M. A., Hernández-Díaz, S. and Robins, J. M. (2004)
A structural approach to selection bias.
Epidemiology, 15:615–625.

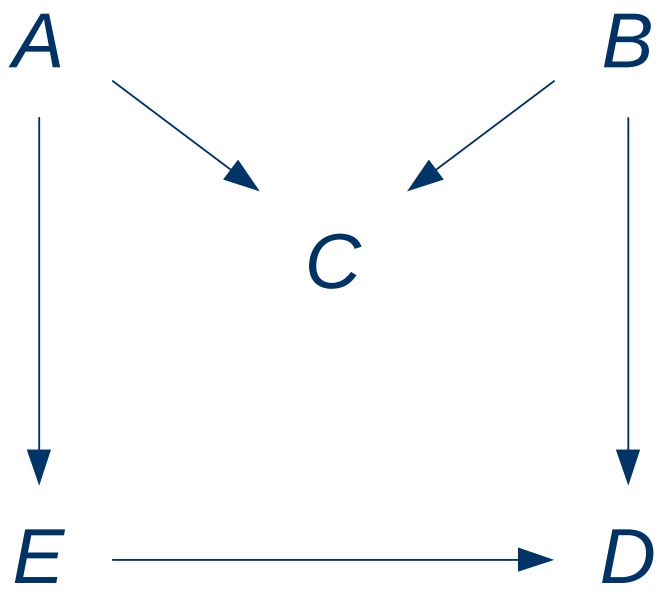
Comparison with the 'traditional' view (1)



What if the confounder doesn't cause the outcome?

- Is C a confounder here?
- Causal diagrams say yes.
- Unclear in the traditional view: if 'risk factors' can be non-causal, then yes.

Comparison with the 'traditional' view (2)



The 'M'-structure

- Is C a confounder here?
- Causal diagram approach says NO. Controlling for it CREATES bias).
- Unclear in the traditional view: if 'risk factors' can be non-causal, then yes.