PATHWAYS

What is the difference between association and causation?

Rhian Daniel and Bianca De Stavola

ESRC Research Methods Festival, 5th July 2012, 10.00am

Website http://pathways.lshtm.ac.uk
Email pathways@lshtm.ac.uk
Twitter @pathwaysNCRM

PATHWAYS

PATHWAYS biosocial influences on health

What is the difference between association and causation?

And why should we bother being formal about it?

Rhian Daniel and Bianca De Stavola

ESRC Research Methods Festival, 5th July 2012, 10.00am

Website http://pathways.lshtm.ac.uk
Email pathways@lshtm.ac.uk
Twitter @pathwaysNCRM

UNIVERSITY OF CAMBRIDGE Department of Geography

NCRM National Centre for Research Methods

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Outline

Outline

What is causal inference? (1)

- Causal inference is the science (sometimes art?) of inferring the presence and magnitude of cause–effect relationships from data.
- As sociologists, economists, epidemiologists *etc.*, and indeed as human beings, it is something we know an awful lot about.
- Suppose a study finds an association between paternal silk tie ownership and infant mortality.
- On the back of this, the government implements a programme in which 5 silk ties are given to all men aged 18–45 with a view to reducing infant mortality.
- We would all agree that this is madness.
- This is because we understand the difference between association and causation.

What is causal inference? (2)

- Much of our research is about cause–effect relationships.
- If we can find modifiable causes of adverse outcomes, we can change the world!
- Modifying factors that are non-causally associated with adverse outcomes is an expensive waste of time.
- The field of causal inference consists of (at least) three parts:
  1. A formal language for unambiguously defining causal concepts. This is just a formalisation of the common sense we already have.
  2. Causal diagrams: a tool for clearly displaying our causal assumptions. They can be used to inform both the design and analysis of observational studies.
  3. Analysis methods (i.e. statistical methods) that can help us draw more reliable causal conclusions from the data at hand.
- In this talk, I will mostly focus on 1 and 2, and briefly mention 3.

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.
- 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).
- 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 $Z = 1$:

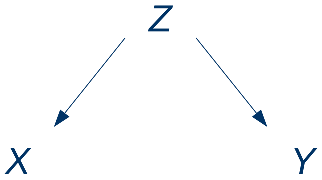
$$P(Y = 1 \,|X = 1) = \frac{1}{3}$$

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

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

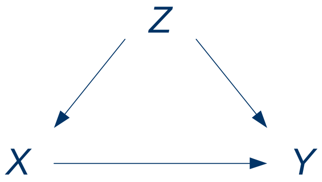
Summary so far (1)

- 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

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

$$X \longrightarrow Y$$

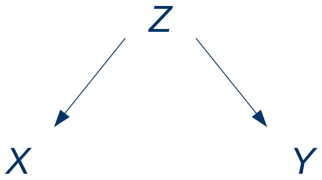
- Two variables X and Y will be associated in the population if X causes Y.

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

$$X \longleftarrow Y$$

- 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)

$$X \longrightarrow Y \qquad X \longleftarrow Y \qquad \begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X & & Y \end{tikzcd}$$

- 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

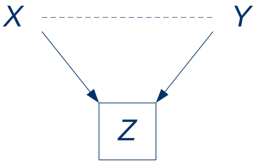
$$P(X = 1 | Y = 1) \neq P(X = 1 | Y = 0)$$

  and

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

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

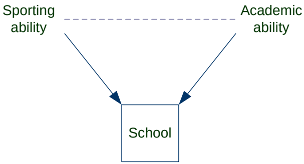
How can two variables be associated in a sub-population? (1)



- 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.
- X and Y will be conditionally associated (given Z), even if they are marginally independent.
- 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



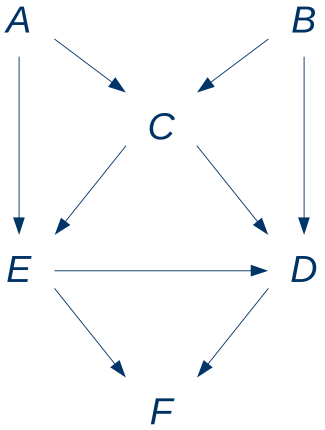
- Suppose there is a selective school that accepts pupils who are either good at sport, or good academically, or both.
- Suppose too that sporting ability and academic ability are independent in the population.
- Within this school, there will be a (negative) association between sporting and academic 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

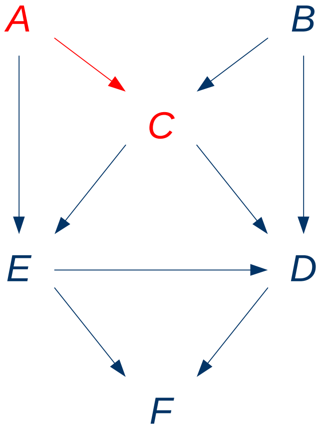
An example



Directed acyclic graph

- This is an example of a causal diagram or causal directed acyclic graph (DAG).
- It is directed since each edge is a single-headed arrow.
- It is causal since the arrows represent our assumptions about the direction of causal influence.
- It is acyclic since it contains no cycles: no variable causes itself. [NB 'Feedback' can be dealt with by incorporating time].

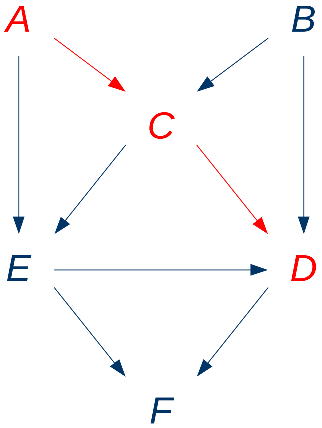
Terminology (1)



Parents and children

- A is a parent of C.
- C is a child of A.

Terminology (2)



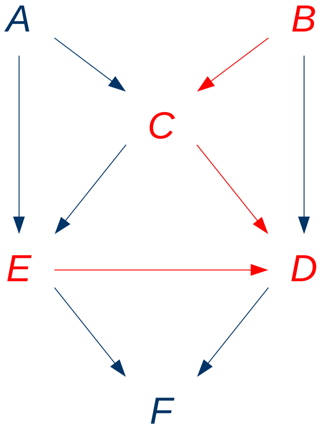
Ancestors and descendants

- A is an ancestor of D.
- D is a descendant of A.

[NB:

- A is also an ancestor of C.
- C is also a descendant of A.
- i.e. parents are ancestors, and children are descendants.]

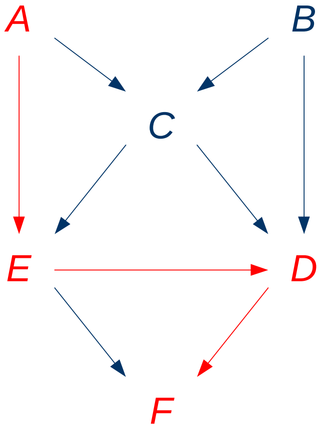
Terminology (3)



Path

- This is a path from E to B.

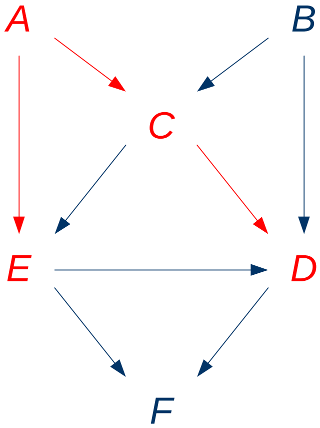
Terminology (4)



Directed path

- This is a directed path from A to F (since all arrows point 'forwards').

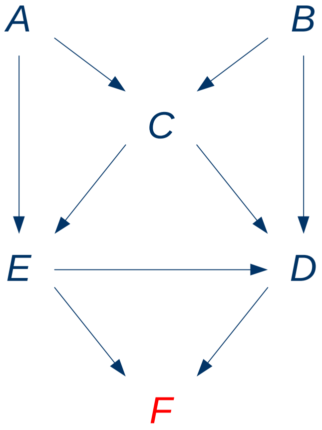
Terminology (5)



Back-door path

- This is a back-door path from E to D, since it starts with an arrow into E.

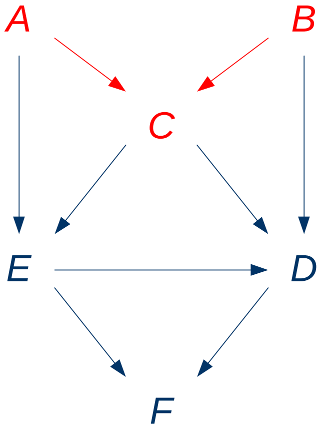
Terminology (6)



Collider

- F is a collider since two arrow-heads meet at F.

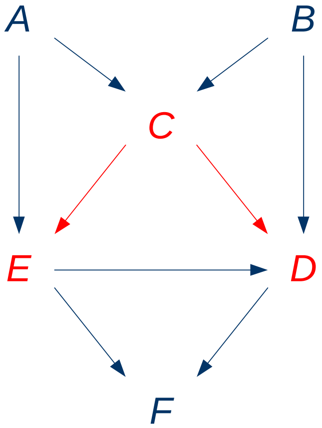
Terminology (7)



Note

- Note that C is a collider on the path $A \rightarrow C \leftarrow B$ ...

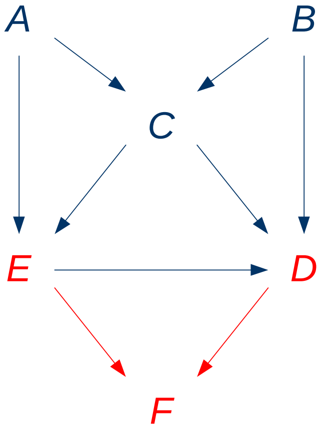
Terminology (8)



Note

- but C is NOT a collider on the path $E \leftarrow C \rightarrow D$.
- Thus the definition of a collider is with respect to the path being considered.

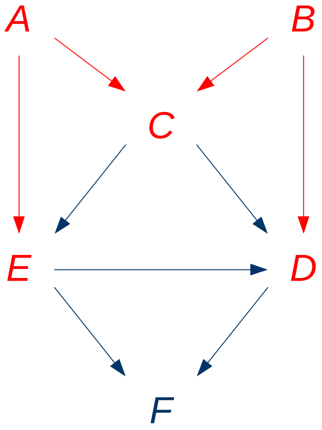
Terminology (9)



Blocked path

- The path $E \rightarrow F \leftarrow D$ is blocked since it contains a collider (F).

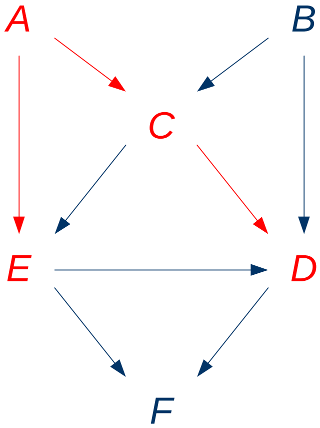
Terminology (10)



Blocked path

- This path is also blocked (at C).

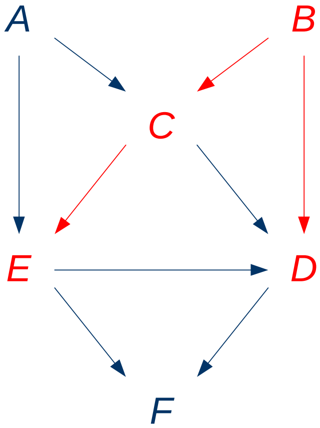
Terminology (11)



Open path

- A path which does not contain a collider is open. Here is an example. . .

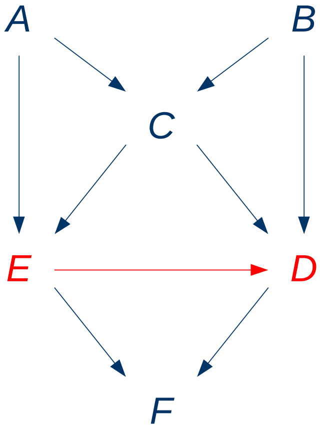
Terminology (12)



Open path

- . . . and another . . .

Terminology (13)



Open path

- . . . and another.

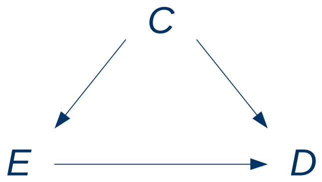
How to construct a causal diagram (1)

$$E \longrightarrow D$$

Step 1

- The first step in constructing a causal diagram for a particular problem is to write down the exposure and outcome (e.g. disease) of interest, with an arrow from the exposure to the outcome.
- This arrow represents the causal effect we aim to estimate.

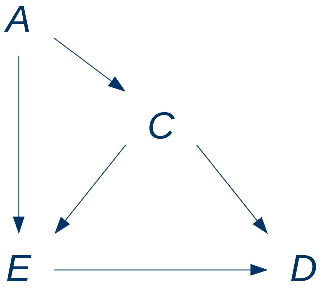
How to construct a causal diagram (2)



Step 2

- If there is any common cause C of E and D, we must write it in the diagram, with arrows from C to E and C to D.
- We must include C in the diagram irrespective of whether or not it has been measured in our study.

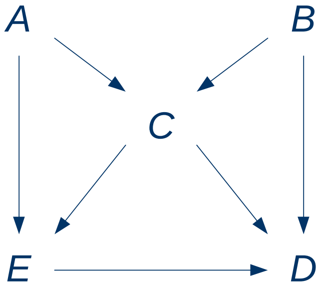
How to construct a causal diagram (3)



Step 2

- We continue in this way, adding to the diagram any variable (observed or unobserved) which is a common cause of two or more variables already in the diagram.

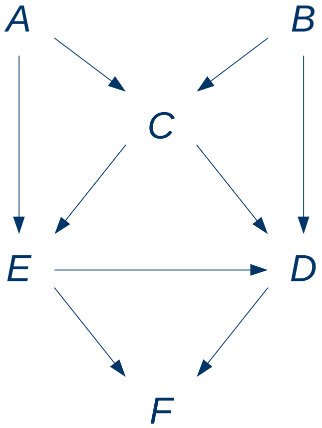
How to construct a causal diagram (4)



Step 2

- We continue in this way, adding to the diagram any variable (observed or unobserved) which is a common cause of two or more variables already in the diagram.

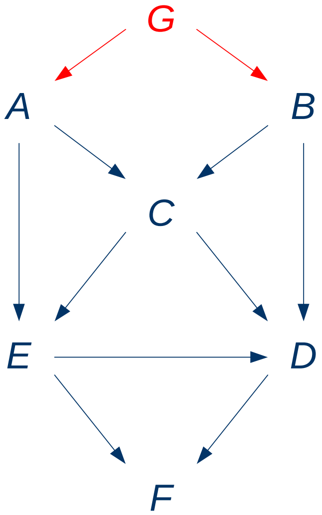
How to construct a causal diagram (5)



Step 3

- If we choose, we can also include other variables, even if they are not common causes of other variables in the diagram.
- For example, F.
- Suppose we finish at this point. The variables and arrows NOT in our diagram represent our causal assumptions.

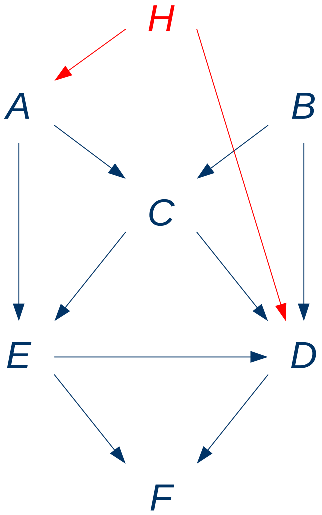
How to construct a causal diagram (6)



What are our assumptions?

- For example, we are making the assumption that there is no common cause G of A and B.

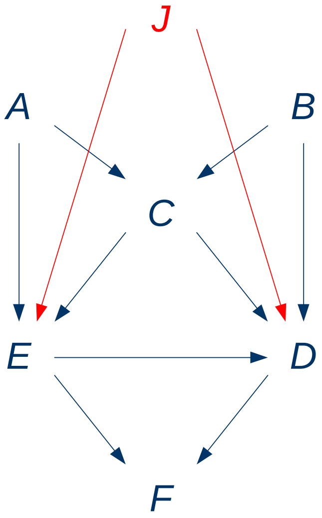
How to construct a causal diagram (7)



What are our assumptions?

- And that there is no common cause H of A and D.

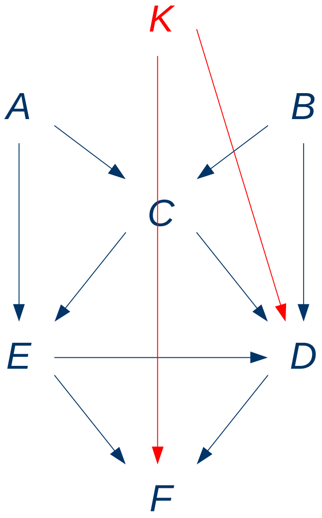
How to construct a causal diagram (8)



What are our assumptions?

- And that A, B and C represent ALL common causes of E and D—there is no additional common cause J.

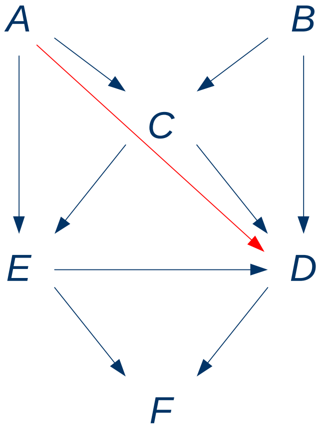
How to construct a causal diagram (9)



What are our assumptions?

- And that there is no additional common cause K of F and D.

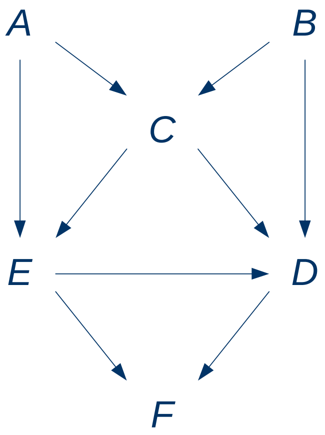
How to construct a causal diagram (10)



What are our assumptions?

- Therefore, each omitted arrow also represents an assumption.
- For example, we are assuming that all the effect of A on D acts through C and E.

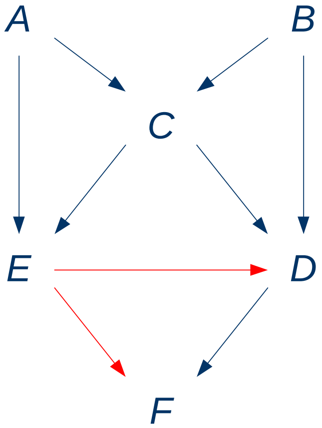
Back-door criterion: is there confounding? (1)



What next?

- IF we believe our causal diagram, we can proceed to determine whether or not the $E \rightarrow D$ relationship is confounded.
- This is done using the back-door criterion.
- The back-door criterion comes in two halves:
  1. the first half determines whether or not there is confounding
  2. if there is, the second half determines whether or not we can control for it.

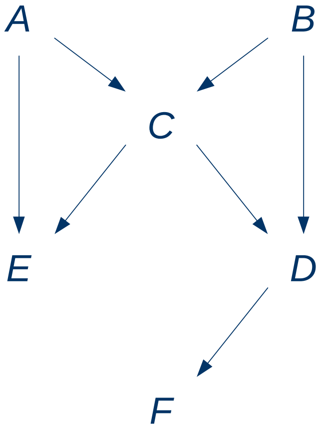
Back-door criterion: is there confounding? (2)



Step 1

- First we remove all arrows emanating from the exposure.

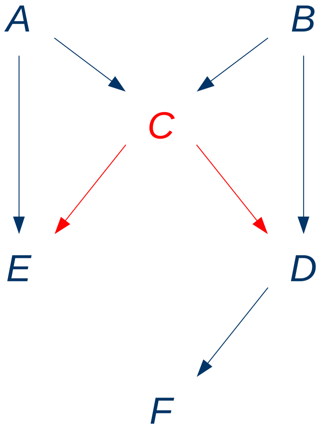
Back-door criterion: is there confounding? (3)



Step 2

- Then we look for any open paths from the exposure to the outcome.
- Recall: an open path does not contain a collider.

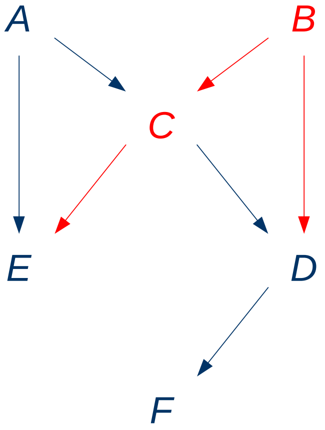
Back-door criterion: is there confounding? (4)



Step 2

- Is this an open path?
- Yes.

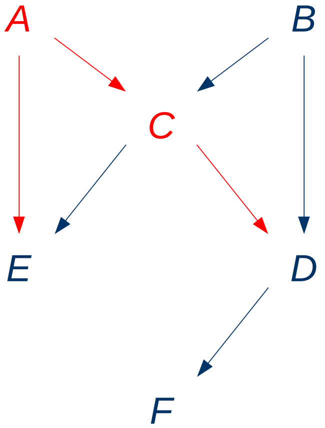
Back-door criterion: is there confounding? (5)



Step 2

- Is this an open path?
- Yes.

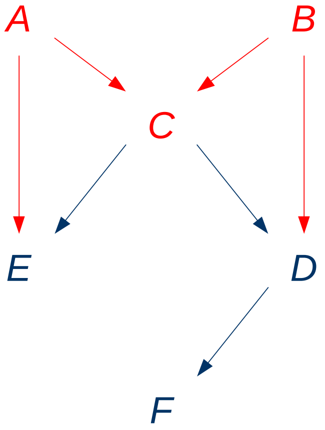
Back-door criterion: is there confounding? (6)



Step 2

- Is this an open path?
- Yes.

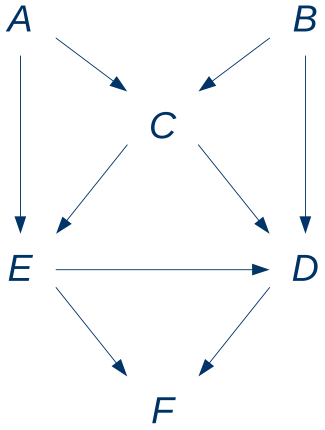
Back-door criterion: is there confounding? (7)



Step 2

- Is this an open path?
- No!

Back-door criterion: is there confounding? (8)



Is there confounding?

- So, we have identified three open back-door paths from E to D. Thus, there is confounding.
- Next question: can we use some or all of A, B, C, F to control for this confounding?
- We have determined that association $\neq$ causation here. But is there a set of variables $\mathcal{S}$ such that if we stratify on them, association $=$ causation within these strata?

The back-door criterion

The second half of the back-door criterion allows us to determine, based on our causal diagram, whether or not a candidate set of covariates is sufficient to control for confounding:

The back-door criterion

(i) First, the candidate set $\mathcal{S}$ must not contain any descendants of the exposure.

(ii) Then, we remove all arrows emanating from the exposure.

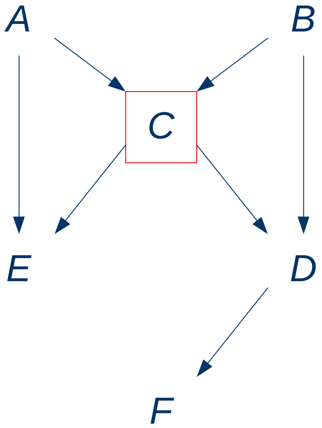
(iii) Then, we join with a dotted line any two variables that share a child which is either itself in $\mathcal{S}$ or has a descendant in $\mathcal{S}$.

(iv) 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.

Let's try this out on our example.

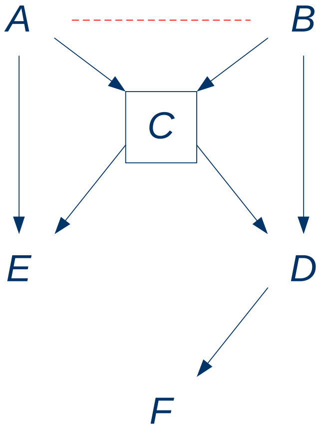
Back-door criterion: can we control it? (1)



The back-door criterion: steps (i) and (ii)

- Is C sufficient?
- C is not a descendant of E, so step (i) is satisfied.
- We have already removed all arrows emanating from the exposure (step (ii)).

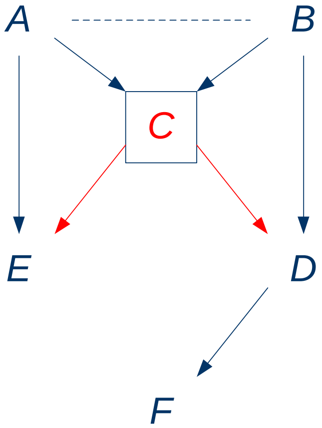
Back-door criterion: can we control it? (2)



Step (iii)

- We join A and B with a dotted line, since they share a child (C) which is in our candidate set (C).
- No other two variables need be joined in this way.

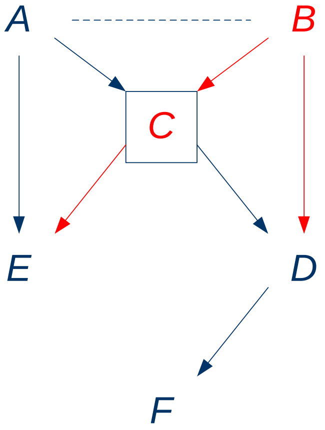
Back-door criterion: can we control it? (3)



Step (iv)

- Now we look for open paths from E to D and see if they all pass through C.
- This one is OK.

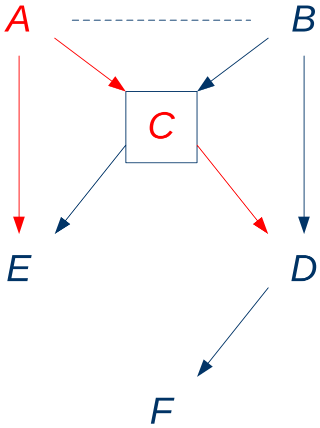
Back-door criterion: can we control it? (4)



Step (iv)

- So is this one.

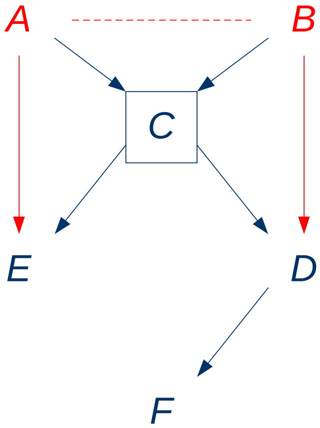
Back-door criterion: can we control it? (5)



Step (iv)

- So is this one.

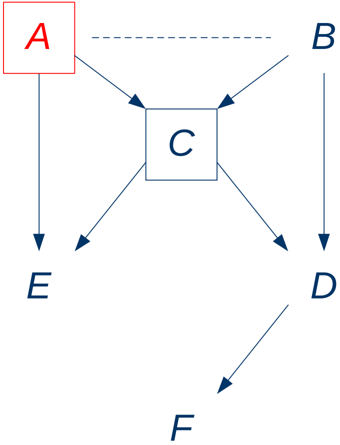
Back-door criterion: can we control it? (6)



Step (iv)

- BUT, here is an open path from E to D that does NOT pass through C.
- So, controlling for C alone is NOT sufficient.

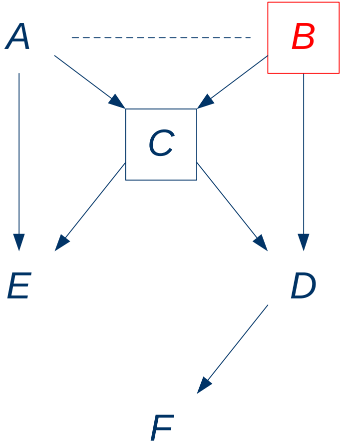
Back-door criterion: can we control it? (7)



What's the solution?

- We must additionally control for either A. . .

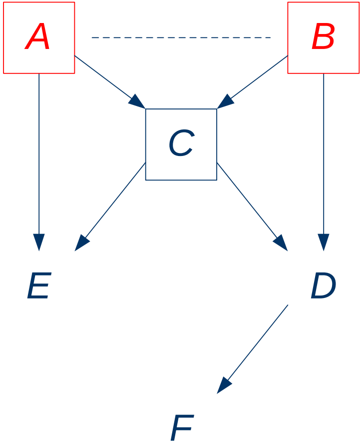
Back-door criterion: can we control it? (8)



What's the solution?

- . . . or B . . .

Back-door criterion: can we control it? (9)



What's the solution?

- ...or both A and B to control for the confounding.

Outline

Why bother?

What has causal inference research (since Rubin 1978) given us? (1)

1. A formal language (counterfactuals, hypothetical interventions) so that age-old causal concepts can be nailed down mathematically, eg
   - causal effect
   - direct effect
   - indirect effect
   - confounding
   - selection bias
   - effect modification
2. Tools for making explicit the assumptions under which our analysis (eg regression) gives estimates that can be interpreted causally, eg
   - causal diagrams (DAGs)

Why bother?

What has causal inference research (since Rubin 1978) given us? (2)

3. When the assumptions needed for ‘standard’ analyses to be causally-interpretable are too far-fetched, alternative methods have been proposed that give causally-interpretable estimates under a weaker set of assumptions, eg (for problems of intermediate confounding)
   - g-computation formula
   - inverse probability weighting of marginal structural models
   - g-estimation of structural nested models

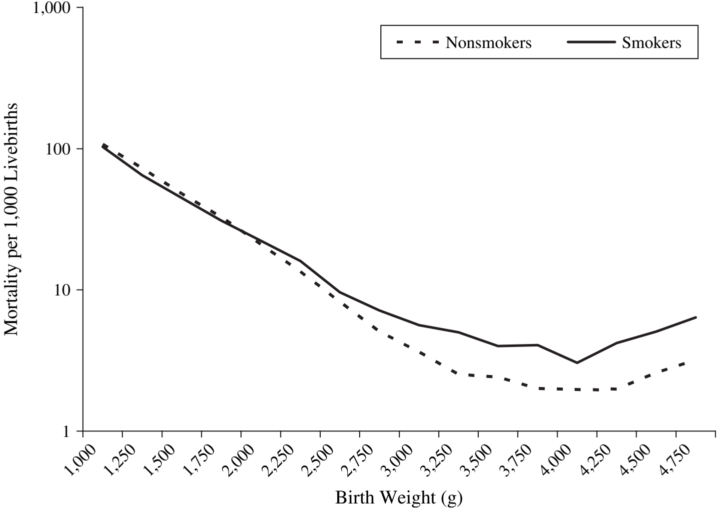
   [Would this have been possible without 1 & 2?]
4. Sensitivity analyses can be performed to see how robust our (causal) conclusions are to violations of these assumptions
   [Not possible without explicit assumptions]

Outline

Example: the birthweight “paradox” (1)

- Many epidemiological studies from the 1960s onwards found that low birthweight (LBW) infants have lower infant mortality in groups in which LBW is most frequent.
- “The increase in the incidence of LBW among infants of smoking mothers was confirmed. However, a number of paradoxical findings were observed which raise doubts as to causation. Thus, no increase in neonatal mortality was noted. Rather, the neonatal mortality rate and the risk of congenital anomalies of LBW infants were considerably lower for smoking than for nonsmoking mothers. These favourable results cannot be explained by differences in gestational age...” (Yerushalmy, AJE 1971)

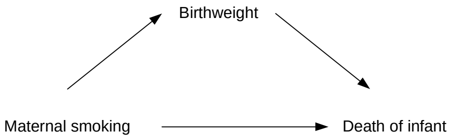
Example: the birthweight “paradox” (2)



Example: the birthweight “paradox”

A ‘causal inference’ view (1)

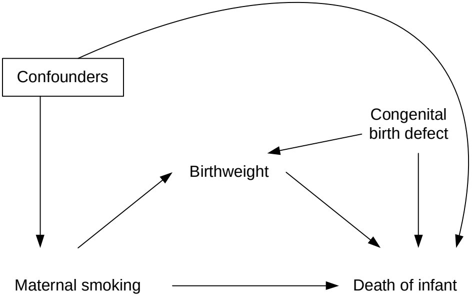
- Hernández-Díaz et al (AJE, 2006) explained this “paradox” using simple causal thinking.



- Birthweight is on the causal pathway from maternal smoking to the death of the child.
- If we wanted the total causal effect of maternal smoking on infant mortality, we shouldn't adjust for BW.
- By adjusting, we are trying to estimate a direct effect. (Point 1).

Example: the birthweight "paradox"

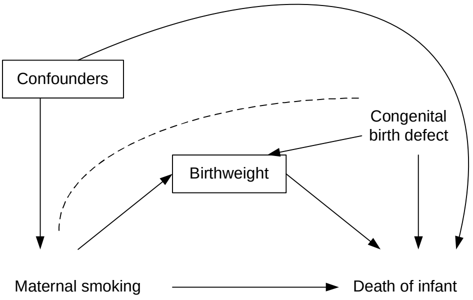
A 'causal inference' view (2)



- But there are common causes of LBW and infant mortality, eg congenital birth defects, and confounders of smoking and infant mortality. (Point 2).

Example: the birthweight "paradox"

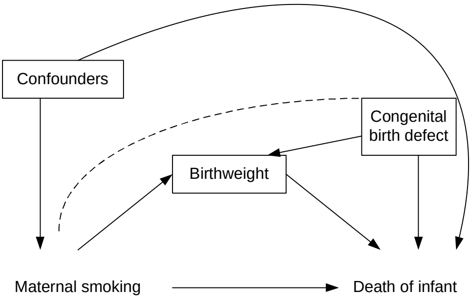
A 'causal inference' view (3)



- Stratifying on the common effect of two independent causes induces an association between the causes. (Why?)
- Congenital birth defects plays the role of a confounder in this analysis.
- This explains the "paradoxical" findings.

Example: the birthweight "paradox"

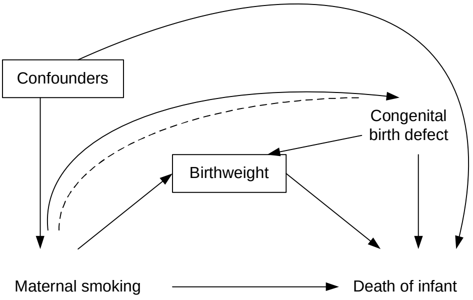
A 'causal inference' view (4)



- So we should adjust for it when looking within strata of birthweight. (Still point 2).

Example: the birthweight "paradox"

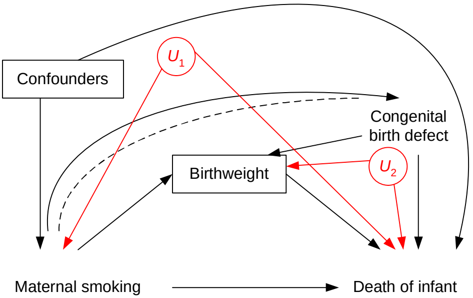
A 'causal inference' view (5)



- But what if maternal smoking also causes congenital birth defects?
- Now it is an intermediate confounder.
- Alternative methods (g-computation, ipw, g-estimation) can be used. (Point 3).

Example: the birthweight "paradox"

A 'causal inference' view (6)



- And what if there are other (unmeasured) common causes of birthweight and infant mortality?
- Sensitivity analyses. (Point 4).

Outline

Why bother?

In conclusion. . .

- If we know the language of causal inference, we are able to:
  - know exactly what we mean when talking about causal effect/direct effect/confounding etc
  - be honest about the assumptions under which association=causation
  - try to use analyses based on more plausible assumptions
  - report how sensitive our causal conclusions are to these assumptions
- Without the language of causal inference, we risk:
  - getting into a muddle when talking about causal concepts
  - sticking to analyses which can be causally-interpretable only under highly implausible assumptions
  - that people will interpret our estimates causally even when we warn them that association≠causation

Final thought

- Always saying "...but association is not causation" is like putting "this product may contain nuts" on all food packaging.
- It's true and absolves us of all responsibility.
- But is it useful? Is it ethical?