What is confounding ……. and what can we do about it?

Roseanne McNamee

Centre for Biostatistics,

Institution of Population Health, University of Manchester, UK

OUTLINE

- What is confounding?
- Association, causes, causal studies, causal graphs
- What is a confounder? … and what it isn't
- What can you do to prevent/reduce confounding?
  - -by study design
  - -by statistical analysis
- Some fundamental problems & more difficult issues

Confounding

- One of the most important issues when considering the validity of observational research concerned with causes

- Examples of causal questions from epidemiology:

  Does HRT have a causal effect on cardiovascular risk?

  Is the MMR vaccine a cause of autism?

  Is (low) birth weight a cause of learning disability?

  Is shift work a cause of heart disease?

In general: is E a cause of D?

where D is the outcome of interest

E is the factor/exposure under investigation

Confounding

is an attribute of a particular study of the E-D relationship

Definitions:

(i) **Confounding** is due to *a lack of comparability* between the Exposed and unexposed groups…

because their disease risks would have been different even if no exposure took place.

(ii**) Confounding** is a situation in which a measure of the effect of an exposure, E, on disease risk is **distorted/biased …**

because of the association of E with other factor(s) that nfluence risk

How big a distortion?

Confounding can

- cause a completely false association

  ie **in truth**, there is no causal relationship between E and D but there is an association in our data.

- can hide a true causal association

  ie **in truth,** there is a causal effect but there is no association in the data

- In extreme cases, can produce an association which is in the opposite direction to the truth.

But sometimes effects less dramatic: the measured associations are s*lightly bigger (smaller)* than the true casual relationships.

Where/when should we worry about it?

Confounding is a **causal concept**: if you are not asking a question about cause, then no need to worry.

Therefore classification of your study objective/question is useful:

- Causal study: is E a cause of Y?
- Descriptive study: how does Y vary across areas of the UK?
- Predictive study: can we find a way of predicting Y from a set of variables X_1 X_2, X_6?

Relative risk

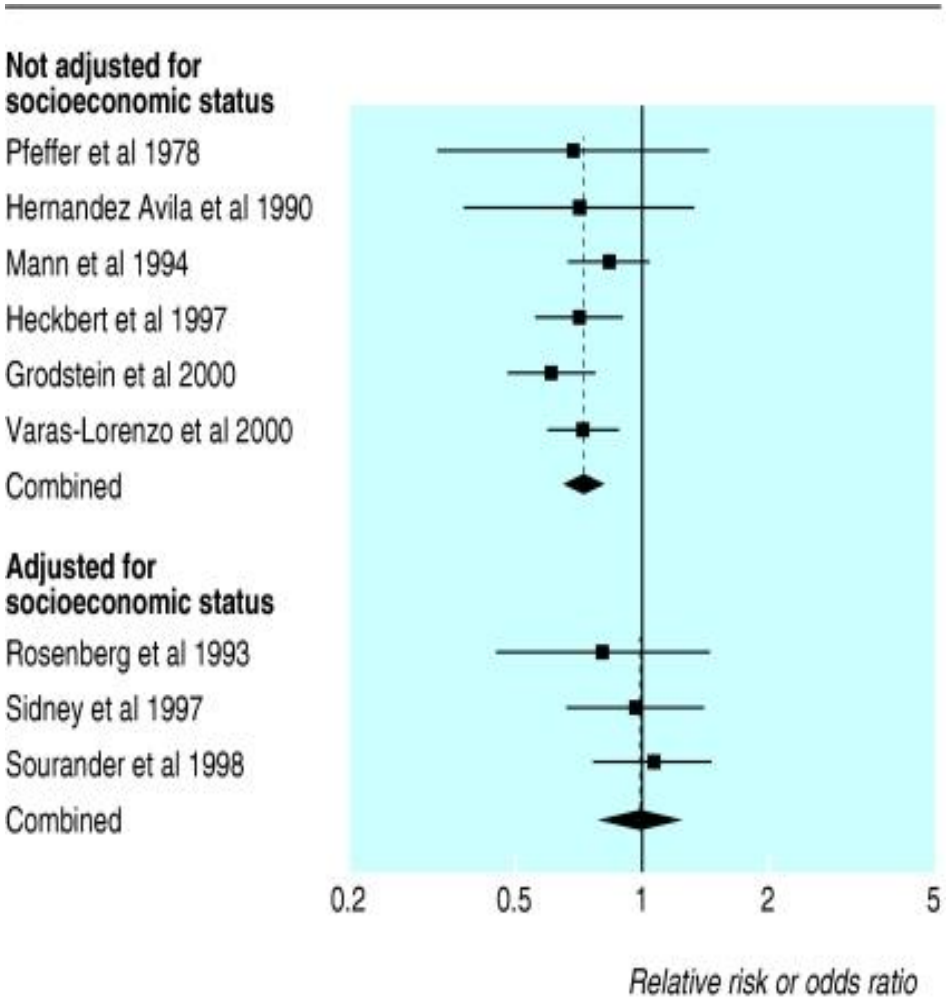
a measure of association between E=exp & D=disease risk

Suppose E is dichotomous

$$\text{Relative Risk(RR)} = \frac{\text{Risk of D in Exposed (E=1)}}{\text{Risk of D in Unexposed(E=0)}}$$

- RR=1: *suggests* no effect of E on D
- RR> 1 ……..E increases risk
- RR< 1 ……..E decreases risk

HRT & cardiovascular disease: what is the true RR?



- Taken from BMJ 2004 (permission to reproduce requested)

The search for confounders….

HRT-CVD relationship:

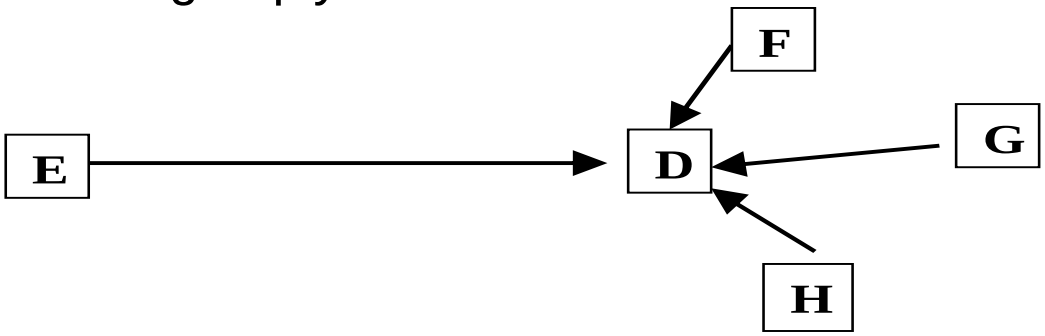
is socio-economic status (SES) a ‘confounding factor’ in the unadjusted analysis?

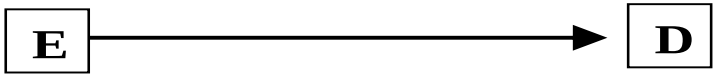
- Confounders = factors which are *jointly* responsible for the distorted measure of the E-D relationship.
- How do we identify confounding factors?
- Adjusted’ analyses: the idea that we can, perhaps, remove the confounding bias by a statistical method.

Causal ‘graphs’

A pictorial method of showing our beliefs about causes…….
& helping us to identify confounders

All the following imply **E is a cause of D:**



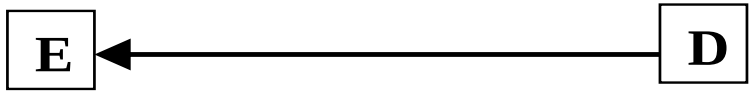


The last diagram will be used as a shorthand

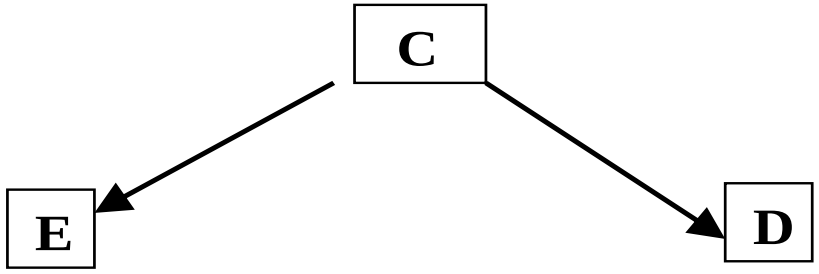
In what situations will we see an association between E and D in a crude data analysis?

(i) **E is a cause of D (see above)**

(ii) **D is a cause of E**



(iii) D and E have at least one cause in common. Here there is one common cause, C:



(iv) We can also have (iii) with (i) or (iii) with (ii)

False association between E and D induced by C – and a solution

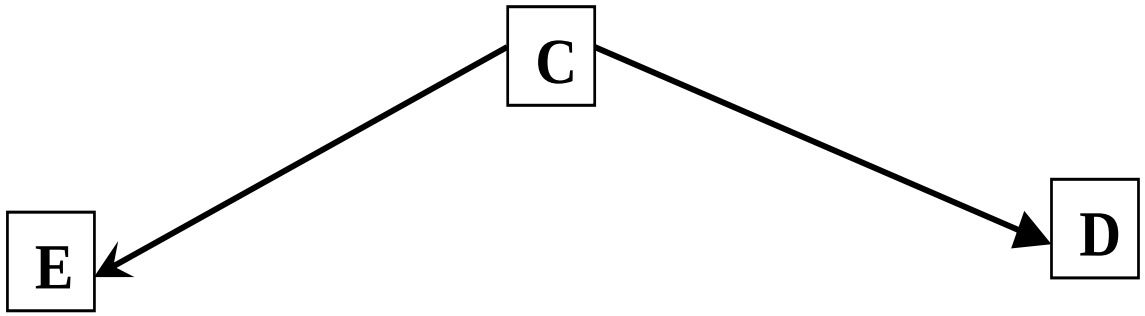


Figure 1

- In truth, NO causal relationship between E and D.
- The relationships C→E, C→D together will induce an *association* between E and D in a crude data analysis
- C is a confounder in a crude analysis of E-D relationship
- Solution: to undo/reduce the bias, adjust for C

False assoc. between E and D induced by relationships, not all of which measured…..

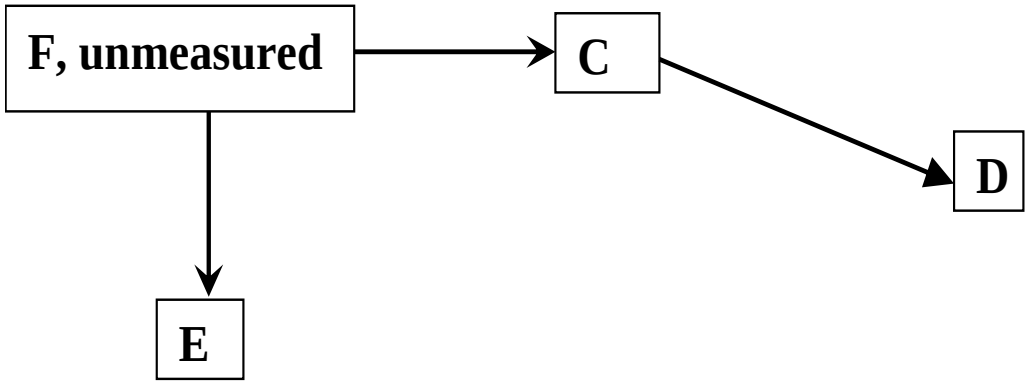


Figure 2

Example: E is alcohol consumption, C = cigarette consumption.

F might be a personal trait which tends to influence smoking and alcohol behaviours

F, the common cause of E and D, is unmeasured.

However: Adjustment for C can undo the confounding.

Adjustment for C is like placing a ‘stopcock’ at C – which breaks the path between E and D.

Conditions (ABC) for a single variable to be treated as a confounder of E-D relationship

A. C must be a cause of D

AND

B C must be correlated with E in the study dataset.

AND

C. C is **not** caused by E - see below where it is

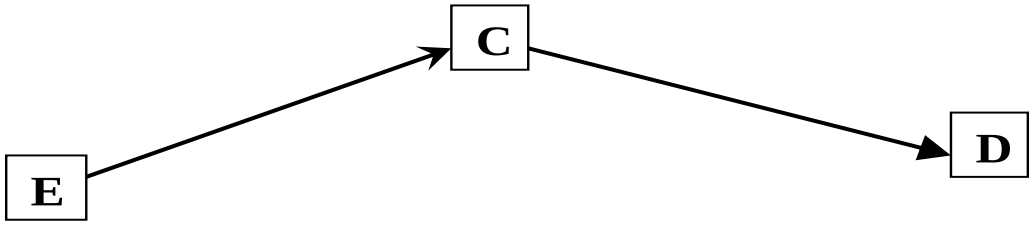


Figure 3: C is not a confounder

eg E = smoking, C = Blood pressure, D= heart disease.

Joint effect of confounders is what matters

Suppose we have two variables, C_1, C_2 which satisfy the confounder conditions.

If the direction of bias (ie distortion) for C_1 alone is opposite in direction to that for C_2, then the joint bias could, in principle be 0!

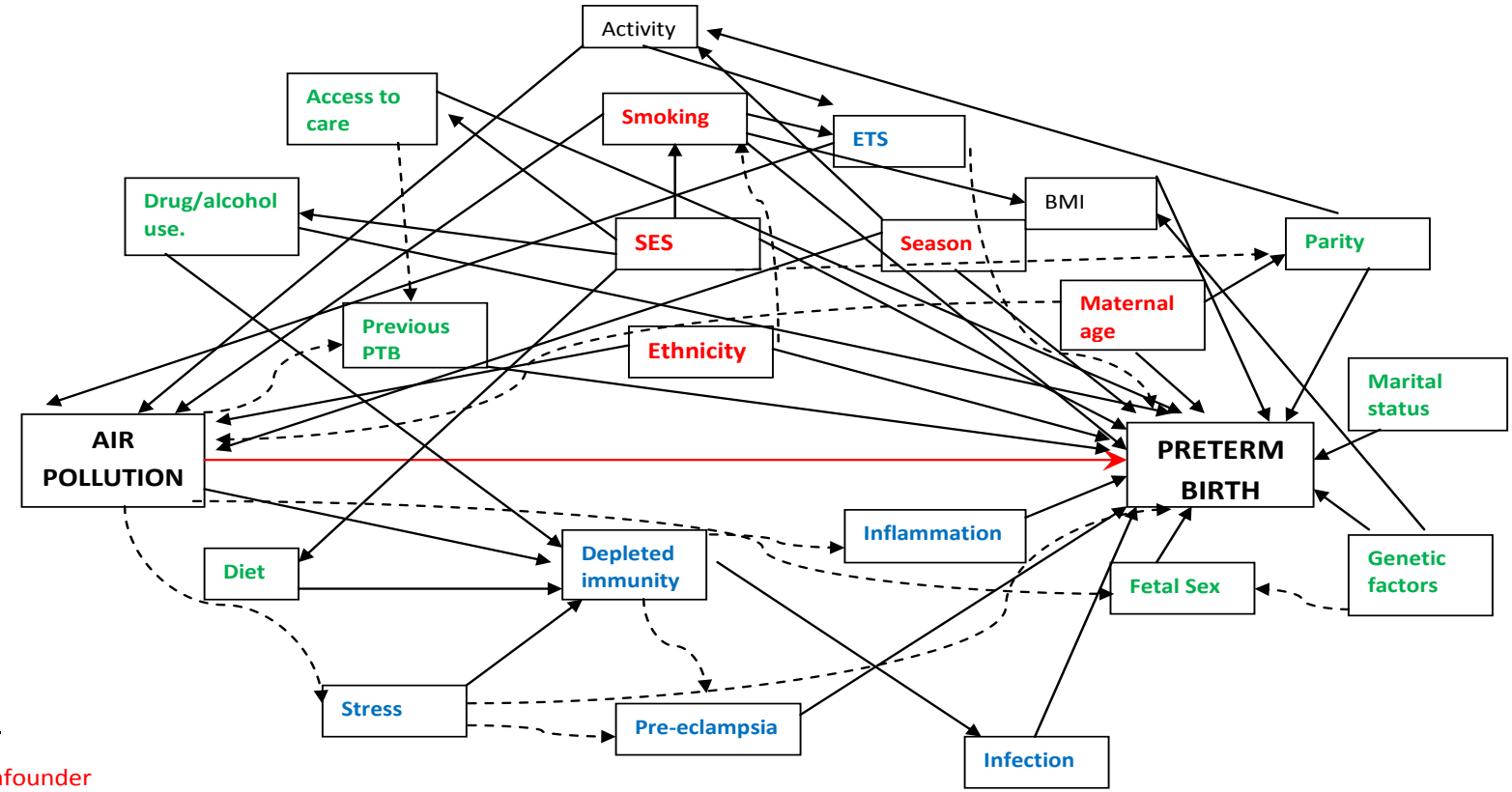
Example (health study comparing Exp and Unexposed):

	Exposed	**Unexposed**
Age	Younger	Older
SES	Higher	Lower

This possibility means that we should be cautious about using previous criteria (ABC) to label variables unanimously as confounders …….

Nevertheless these criteria remain useful …..

More complex causal scenarios…….



Key

= Confounder

Smoking = Mediator

Smoking = Risk Factor only

ETS- Environmental tobacco smoke

Thanks to KH for her picture!

The appeal of causal graphs for inferring confounding in complex problems

- Causal graphs* can tell us whether we can remove confounding by adjusting for a given set of variables….

  no calculation involved!

  no statistical knowledge needed!

- Graphs may show that adjustment for a subset of variables is enough …because it blocks the paths of all others

  …..think of the stopcock analogy

- * assuming graph is correct

Dealing with confounding: statistical methods of adjustment

For measured variables C_1 ,C_2 etc

- Via regression models
- Via stratification
- Via propensity scores

For unmeasured confounders

- Instrumental variable (IV) methods: main issue here is whether we can identify a suitable IV….

Reduction of confounding by study design

Recall:

confounding is a lack of comparability between the Exposed and unexposed groups…

because their disease risks would have been different even if no exposure took place.

- Statistical methods try to deal with consequences -retrospectively
- Can we achieve comparability by study design?
- *Meaning of group comparability* : the groups would have the same outcome – on average- if no exposure took place.

Reduction of confounding by study design

The idea is to find E and not-E groups that are ‘comparable’

- **Randomisation:** randomly allocate individuals to E and not-E gps.

- **Restriction**. restrict study to subjects with a particular value of C. eg restrict study to women.

- **Matching**. for each E individual, find a not-E with same C.

Some fundamental problems

- **Unmeasured confounders** in observational studies:

  just because we don't know about them, doesn't mean they don't exist!

  (Only in randomised studies, can we feel more secure).

- **Measurement error** in confounders:

  Imprecise measurement of C means inadequate adjustment: there will still be 'residual confounding' by C

- A decision to treat C as a confounder of an E-D relationship …… (correctly) depends on making causal assumptions about C!

Cautions

- There is no foolproof statistical ‘test ‘ for confounding:
  we should be guided by external knowledge and data, not data alone…..

- It is possible to create confounding by adjusting for the wrong thing…..

- Time dept confounding: standard statistical approaches for dealing with confounders are invalid & special methods needed!

  If your Es and Cs change over time, & you have a longitudinal study, you need to consider this possibility

Some references

Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health 2001.

McNamee R. Confounding and confounders. Occ Environ Med 2003.

Greenland S, Pearl J, Robins J. Causal diagrams for epidemiological research Epidemiology 1999.

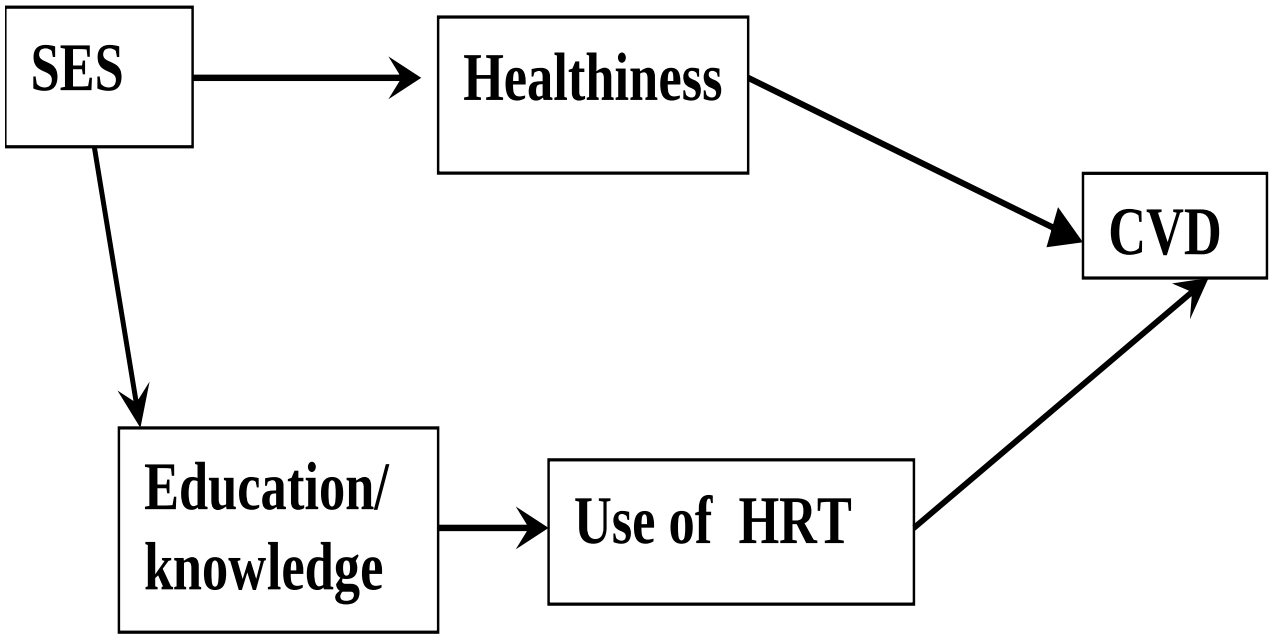
Jewell N. Statistics for epidemiology . Chapman & hall, 2004.

Daniel R, Cousens S, De Stavola B, Kenward M,Sterne J. Methods for dealing with time-dependent confounding. Stat Med Dec 2012.

Also see

http://www.population-health.manchester.ac.uk/biostatistics/research/causalgroup/

A (hypoth) causal graph for HRT/CVD relationship

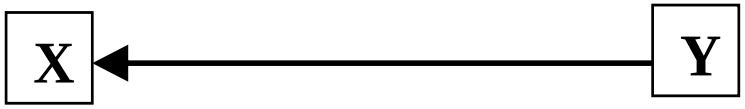


Why do X and Y show an association in a crude data analysis?

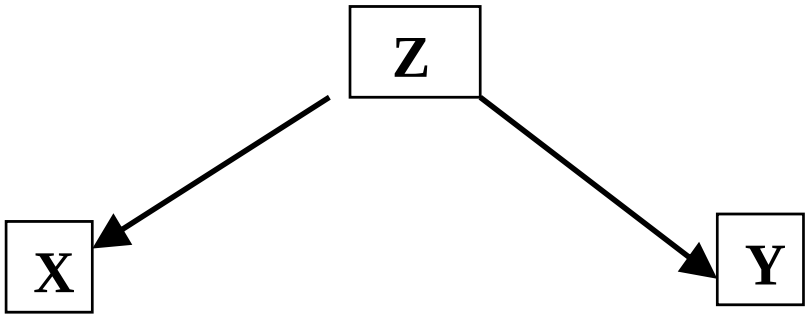
3 reasons:

(i) **X is a cause of Y**

(ii) **Y is a cause of X:**



(iii) X and Y have at least one cause in common. Here there is one common cause, Z:



Association vs causation

Association: a statistical term referring to observed data: it does not carry any causal meaning.

(Measures of association:
correlation coefficients, mean differences , RR, etc)

Example:

RR measured from study data = $RR_{observed}$, say

If $RR_{observed} \neq 1$, there is an association between E & D.

Confounding: $RR_{observed} \neq RR_{true}$