Epidemiology and epidemiological methods

Timothy Mak

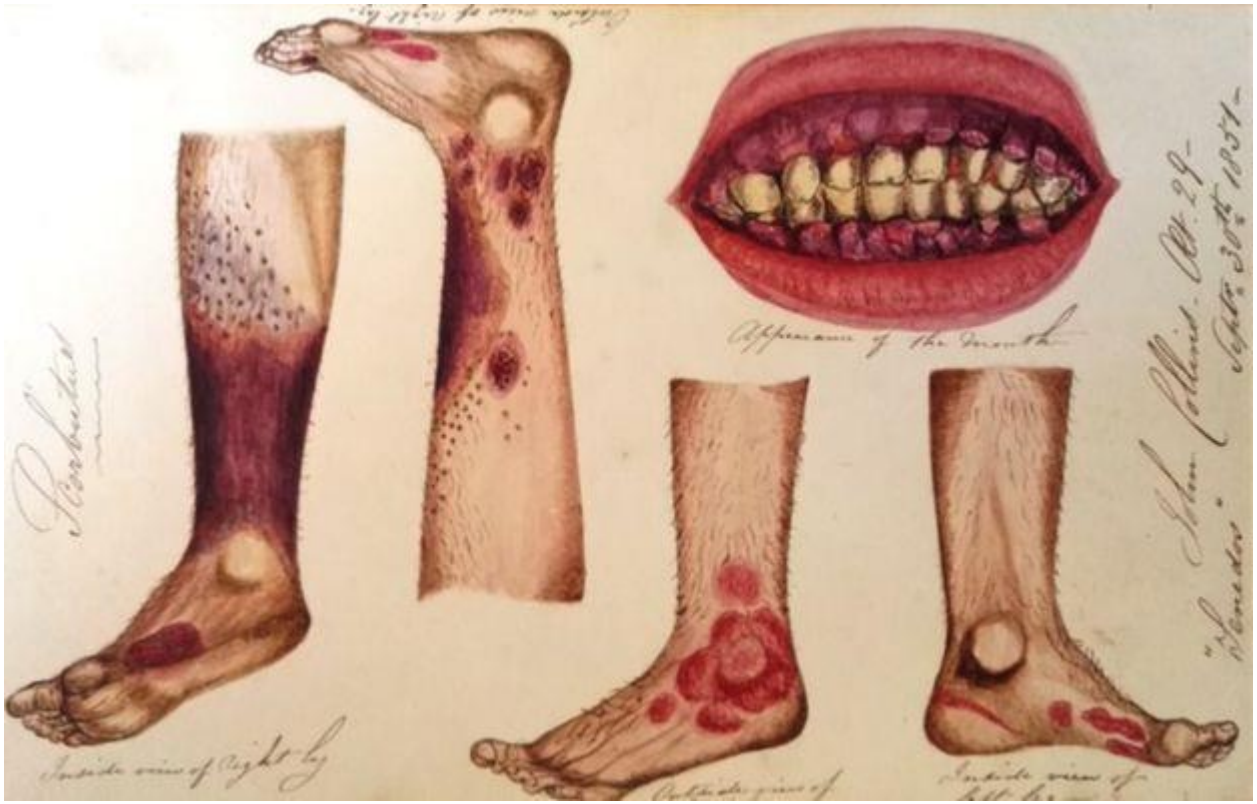
Basic Research Skills seminar
26/6/2016

Epidemiology

"Epidemiology is the study and analysis of the patterns, causes, and effects of health and disease conditions in defined populations." (Wikipedia)

Historical anecdotes

- Scurvy and lemons (James Lind, circa 1747)



Historical anecdotes

- Smallpox and cowpox (Benjamin Jesty, circa 1770)

Modern epidemiology

- Much greater concern with chronic diseases

Cancer epidemiology

How do we know cancer is avoidable? (Doll and Peto, 1981)

- Differences in incidence between communities, after adjustment for age (e.g. lung cancer 35x higher in England among males than in Nigeria, stomach cancer 25x higher in Japan than Uganda)

Cancer epidemiology

How do we know cancer is avoidable? (Doll and Peto, 1981)

- Changes in incidence on migration (comparison of rates in original country, migrant group, and new country)

Cancer epidemiology

How do we know cancer is avoidable? (Doll and Peto, 1981)

- Change in incidences over time (barring changes in detection rates)

Cancer epidemiology

Proportions of cancer deaths attributed to various different factors (Doll and Peto, 1981, Journal of the National Cancer Institute)

Factor or class of factor	Percent of all cancer deaths (best estimate)	Percent of all cancer deaths (range)
Tobacco	30	25-40
Alcohol	3	2-4
Diet	35	10-70
Food additives	<1	-5-2
Reproductive and sexual behaviour	7	1-13
Occupation	4	2-8
Pollution	2	1-5
Industrial products	<1	<1-2
Medicines and medical procedures	1	0.5-3
Geophysical factors	3	2-4
Infection	10?	1-?
Unknown	?	?

(The percentages are not supposed to add up to 100%)

Why Tobacco?

- Lung cancer made up a huge proportion of cancers then (around 1/3 of American male cancer deaths).
- Tobacco smoke associated with hugely increased risk of lung cancers (20x), as well as many other cancers.

EPIDEMIOLOGICAL METHODS

Types of studies

- Randomized controlled trials
- Observational studies:
  - Cohort studies
  - Case-control studies
  - Ecological studies

Randomized controlled trials

- Split into two groups randomly
- Double-blinding
- compare two groups

Observational studies

- Cannot rule out confounding (association not causation)
- Stratification

Stratification

E.g. Effect of coffee consumption on risk of myocardial infarction (heart attack) stratified by smoking status

- Evaluate risks of coffee drinkers vs. non-coffee drinkers among smokers
- Evaluate risks of coffee drinkers vs. non-coffee drinkers among non-smokers
- Combine (average) the two estimates to get population average.
- Test for differences between smokers and non-smokers for effects of coffee consumption on MI. (Interaction effects)

Modeling

- When there are many potential confounders, we cannot stratify by all possible combinations.
- Modeling is required
- Most common is regression analysis

Regression vs. stratification

- Logistic regression with adjustment for confounders is **not** the same as stratification
- But the results are usually approximately equivalent (Greenland and Maldonado, 1994)

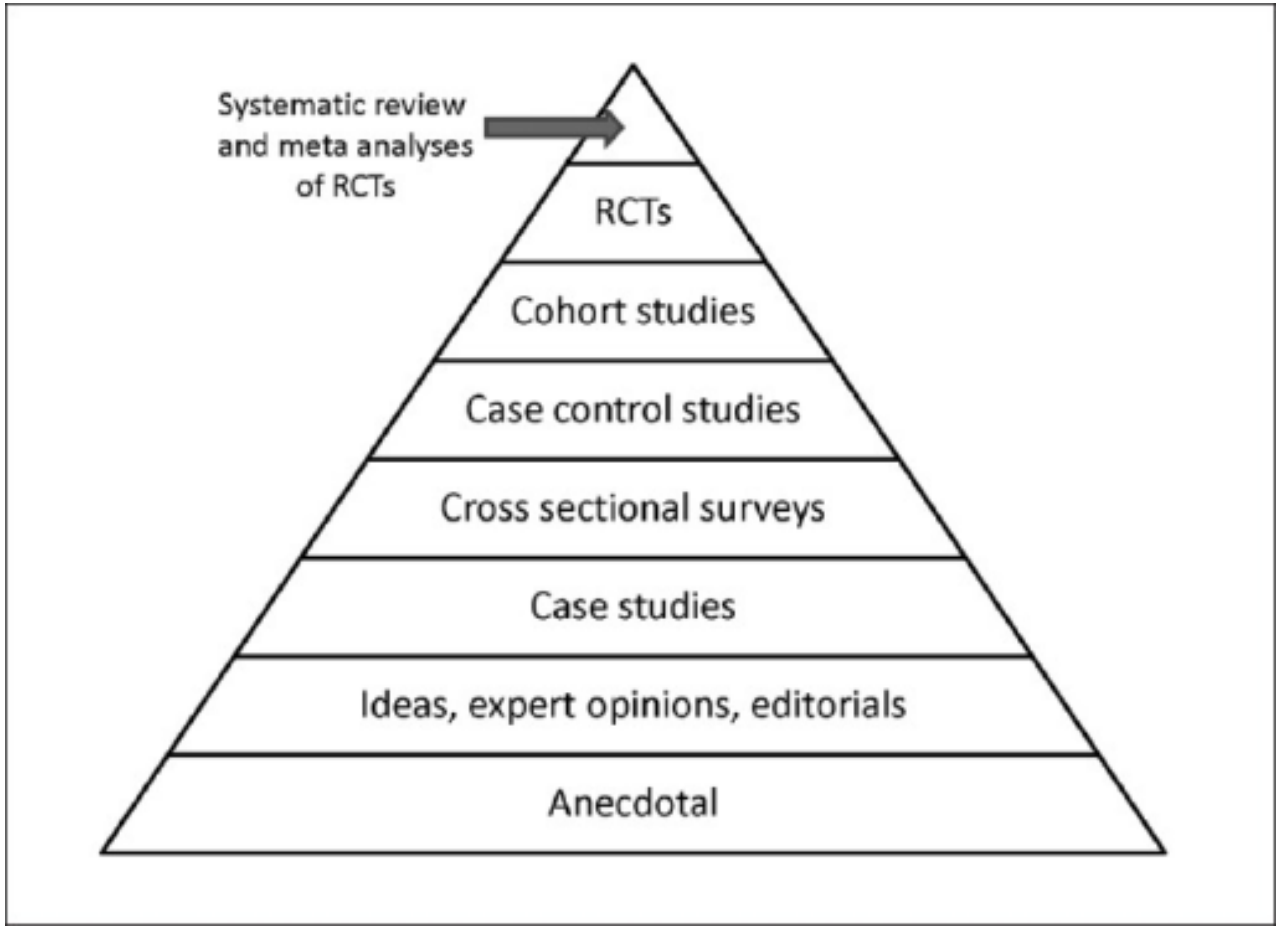
Bias

- Residual confounding
- Measurement error (exposure, outcome, and confounders)
- Selection bias

Bradford Hill criteria (1965)

- Strength of association (effect size)
- Consistency (reproducibility)
- Specificity
- Temporality
- Biological gradient
- Plausibility
- Coherence
- Experiment
- Analogy (with other similar causal relationships)

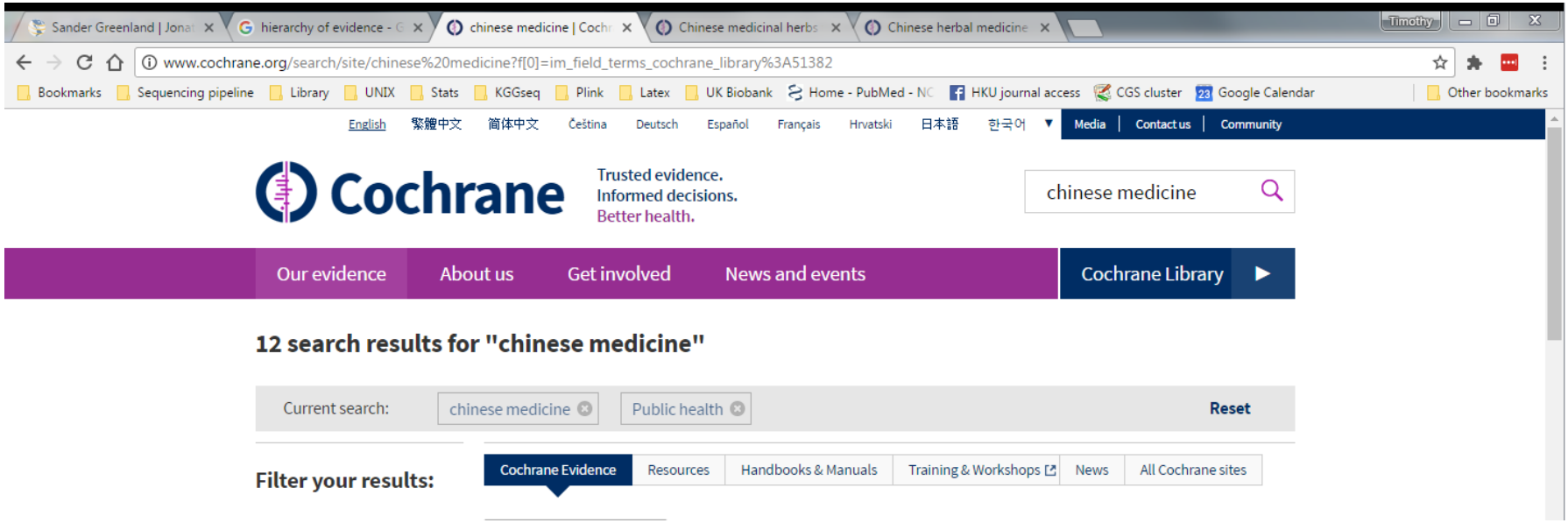
Hierarchy of evidence



https://openi.nlm.nih.gov/imgs/512/154/3168054/PMC3168054_IJSTD-30-112-g001.png

Cochrance reviews (www.cochrane.org)

- Systematic reviews
- systematic assessment of bias
- Continually updated (as far as possible)



ASSESSING POTENTIALS FOR BIAS USING CAUSAL DIAGRAMS

Causal diagrams

- Greenland, Pearl, and Robins (1999) Epidemiology 10(1) 37-48
- Used for determining what and what not to control for in an analysis
- Helps clarifies the various concepts of biases, such as confounding bias, selection bias, measurement error, etc.

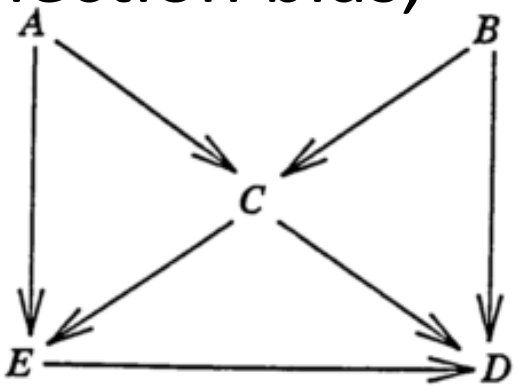


FIGURE 1.

Causal diagrams

- Theoretical model (not based on sample relationships)
- Based on conditional probabilities
- Not necessarily linear
- Directed, Acyclic (Directed Acyclic Graphs)
- Graphs: Made up of nodes and edges

Causal diagrams

- Two points are connected if you can draw a line through them through the edges
- However, a node with two arrows pointing towards it is a collider. A collider blocks a path.
- Here, C is a collider, and the path EACBD is blocked, but ECD is not.

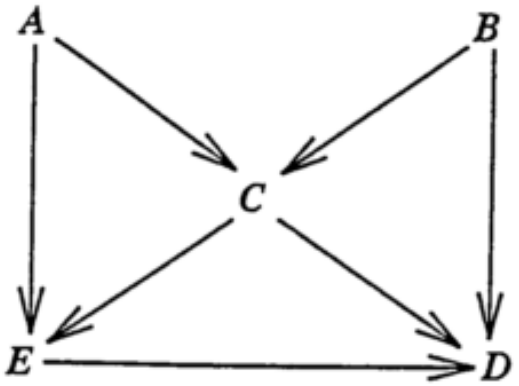


FIGURE 1.

Causal diagrams

- When you control for (condition on / stratify by) a variable (or restrict your analysis to a subgroup in that variable), you break that path.
- e.g. if you condition on B, you break the path CBD.

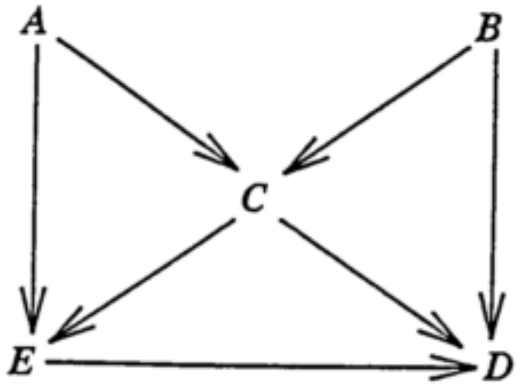


FIGURE 1.

Causal diagrams

- However, when you control for a collider, or the ***descendents*** of a collider, you open up the blocked paths.
- e.g. if you condition on C, EACBD is now unblocked, although ECD is.
- If you condition on D, you also get the path EACB.

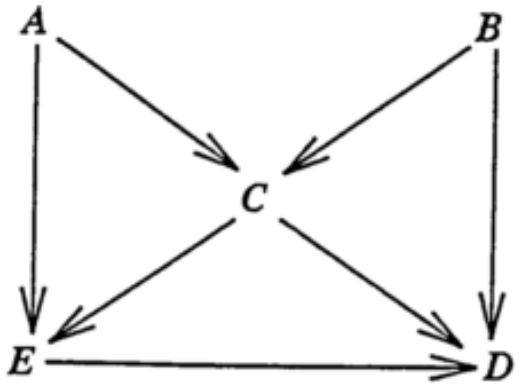
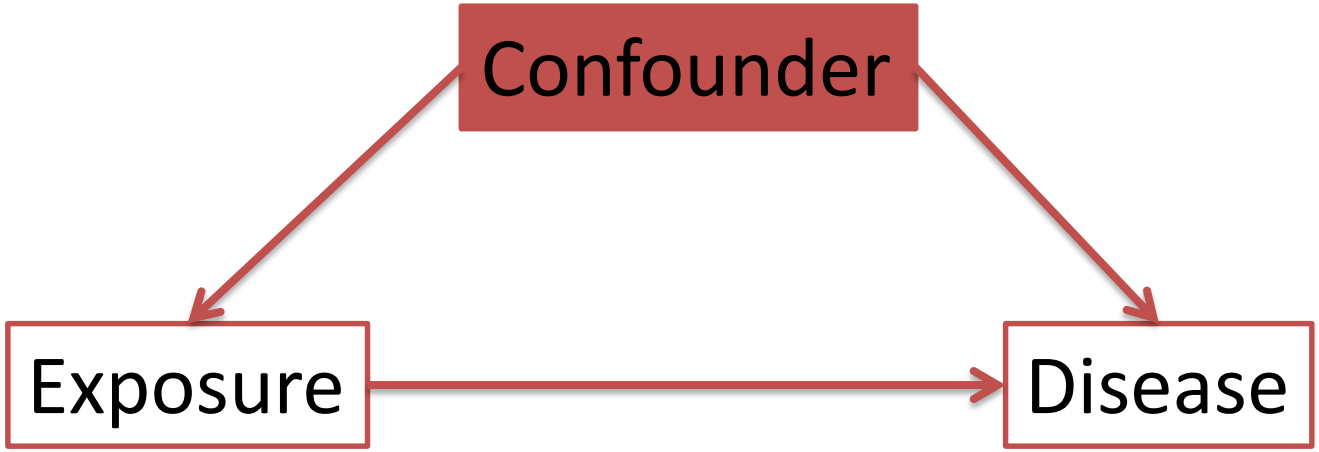
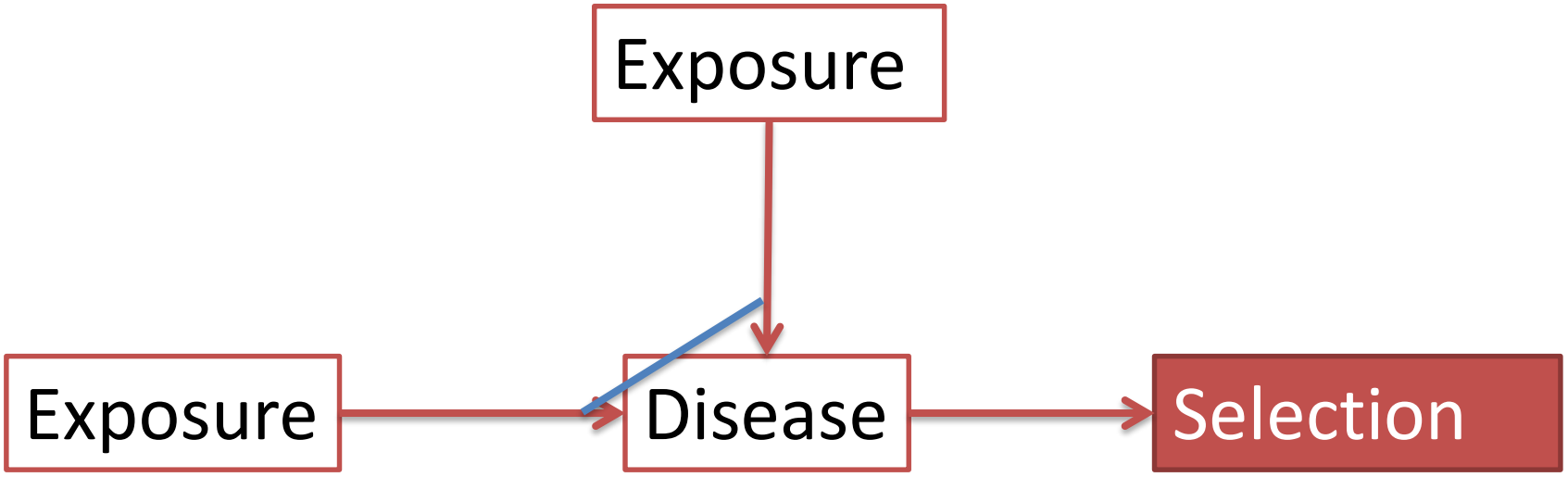


FIGURE 1.

Confounding bias



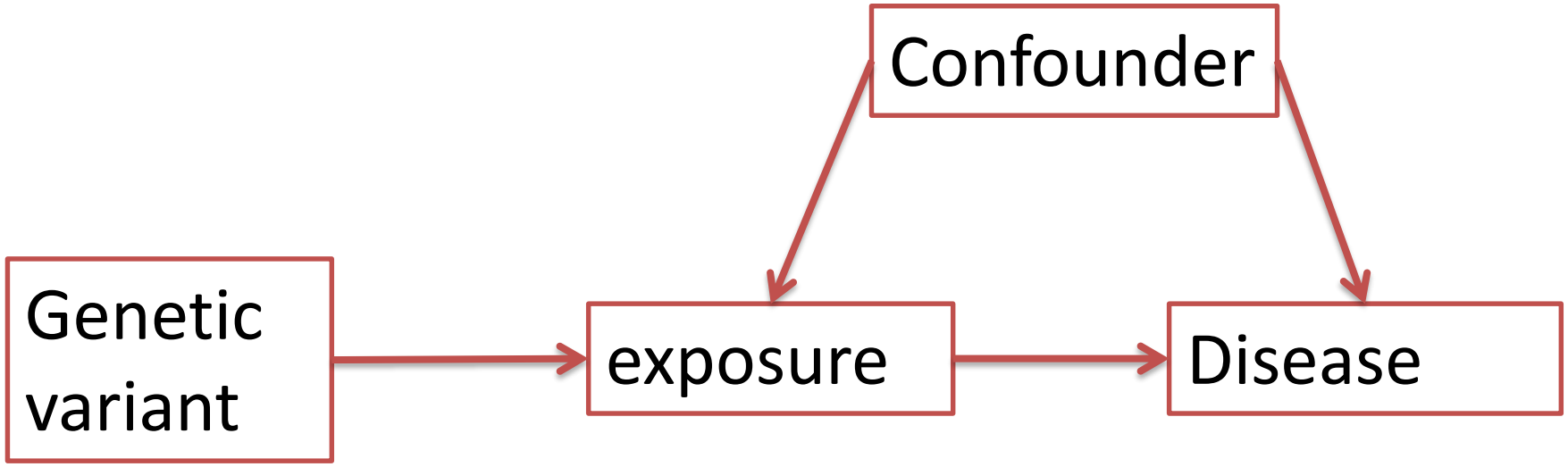
Case-control study



Biased estimates if you examine other relationships in a case-control study rather than with the disease of interest

Measurement error

Mendelian randomization



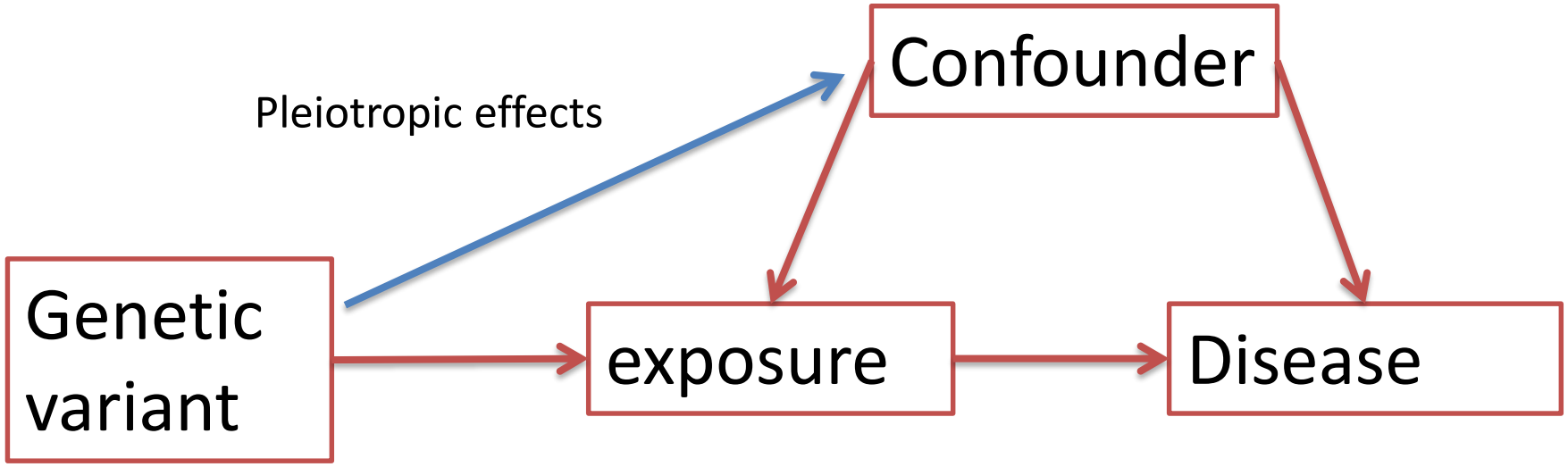
- Exposure is a collider blocking the path from Genetic variant to disease via confounder
- Genetic variant is randomly assigned and hence not confounded.

Mendelian randomization

Why are genetic variants random?

- Meiosis is a random event, but that's conditioned on your parents genotype.
- Are your parents' genotype random? Random mating a crucial assumption
- Ethnicity is a determinant of genetic variant. Hence GWAS are subject to the confounding effects of population stratification.
- Social classes can also affect randomness of mating, but effect on genetic variants small.

Mendelian randomization



- The presence of pleiotropic effects is another reason why effect size from MR can be biased.
- Genetic instruments should be very specific for the exposure.

Mendelian randomization

- Mostly used in testing.
- Test for association between outcome and genetic variant instead of the real exposure of interest.
- However, power is low when the relationship between the variant and the exposure is weak.

Some success stories

- Effect of milk/calcium on Bone mineral density (Genetic variant for lactose intolerance as instrument)
- Vitamin D deficiency and Tuberculosis
- Cholesterol on CHD (variant on hypercholesterolemia as instrument)