What is a case control study?

Tarani Chandola
Social Statistics
University of Manchester

Imagine….

- It is 1950
- You suspect an association b/w smoking and lung cancer
- How would you show this?

Cases

- Find cases of lung cancer (where?)
- What is the exposure?

  Hypothetical scenario:

  - You find 100 cases of lung cancer

  - 60 of them were heavy smokers (>25 cigs/day)

  - Does this show an association between smoking and lung cancer?

Controls

- Where can you find similar groups of patients to lung cancer patients?

Hypothetical scenario:

- You find 100 control patients who do not have lung cancer
- 20 of them were heavy smokers
- So 20% of controls and 60% of cases were heavy smokers
- Does this show an association between smoking and lung cancer?

Comparing cases and controls

- Populations of cases and controls differ
- So cannot compare % who are exposed (different denominators/populations)
- But you can compare the odds of having lung cancer for two groups: the exposed and the unexposed

Doll and Hill (1952)

- Data was collected from hospitalized patients from more than 20 hospitals in London over a four year period (April 1948 and February 1952).
- Investigators asked hospital personnel to contact them whenever a patient was admitted to the hospital with a new diagnosis of lung cancer (cases).
- Investigators also selected a random sample of patients from the same hospitals, but with different illnesses (controls).
- Cases and controls were interviewed about their smoking habits.

Doll and Hill (1952)

- 1465 case-patients were interviewed for the study all under the age of 75, 1357 men and 108 women.
- Because of the disproportionate ratio of men to women (25 to 2), only men were included in the final study.
- An equal number of controls (1357) were interviewed about their smoking history.
- Investigators divided smoking history into groups based on the average number of cigarettes smoked per day.

Doll and Hill (1952)- results

Cigarettes smoked daily	Cases	Controls
0	7	61
10-14	565	706
15-24	445	408
25+	340	182
All Smokers	1,350	1,296
Total	1,357	1,357

What numbers will you compare?

Doll and Hill (1952)- analysis

Cigarettes smoked daily	Cases	Controls	Odds of smoking if case	Odds of smoking if control
0	7	61	=7/1350	=61/1296
10-14	565	706	=565/1350	=706/1296
15-24	445	408	=445/1350	=408/1296
25+	340	182	=340/1350	=182/1296
All Smokers	1350	1296	1	1
Total	1357	1357		

Doll and Hill (1952)- Odds Ratios

Cigarettes smoked daily	Cases	Controls	Odds of smoking if case	Odds of smoking if control	Odds of being a case given exposure
0	7	61	0.01	0.05	=0.02/0.05
10-14	565	706	0.42	0.54	=0.42/0.54
15-24	445	408	0.33	0.31	=0.33/0.31
25+	340	182	0.25	0.14	=0.25/0.14
All Smokers	1350	1296	1.00	1.00	1.00

Doll and Hill (1952)- Odds Ratios

Cigarettes smoked daily	Cases	Controls	Odds of smoking if case	Odds of smoking if control	Odds ratio	Odds ratio (rescaled)
0	7	61	0.01	0.05	0.11	1.00
10-14	565	706	0.42	0.54	0.77	6.97
15-24	445	408	0.33	0.31	1.05	9.50
25+	340	182	0.25	0.14	1.79	16.28

As you move up the groups, from no cigarettes per day to 25+ cigarettes per day, the values of the odds ratio rise steadily, consistent with a dose-response relationship between the exposure (daily number of cigarettes smoked) and the strength of the association (odds ratio).

Odds ratio

- An approximation of relative risk
- RR → “how much more likely are exposed people to get the disease than the unexposed”
- OR → “how much more likely were people with the disease to have been exposed than those without the disease”

Calculation of odds ratio

	Disease + (cases)	**Disease - (controls)**
Exposure +	a	b
Exposure -	c	d

Calculation of odds ratio in 2 x2 table

	Disease + (cases)	**Disease - (controls)**
Exposure +	a	b
Exposure -	c	d

Odds ratio = a:c/b:d = ad/bc

= cross-product ratio

Bias

- What types of bias in a case-control study?

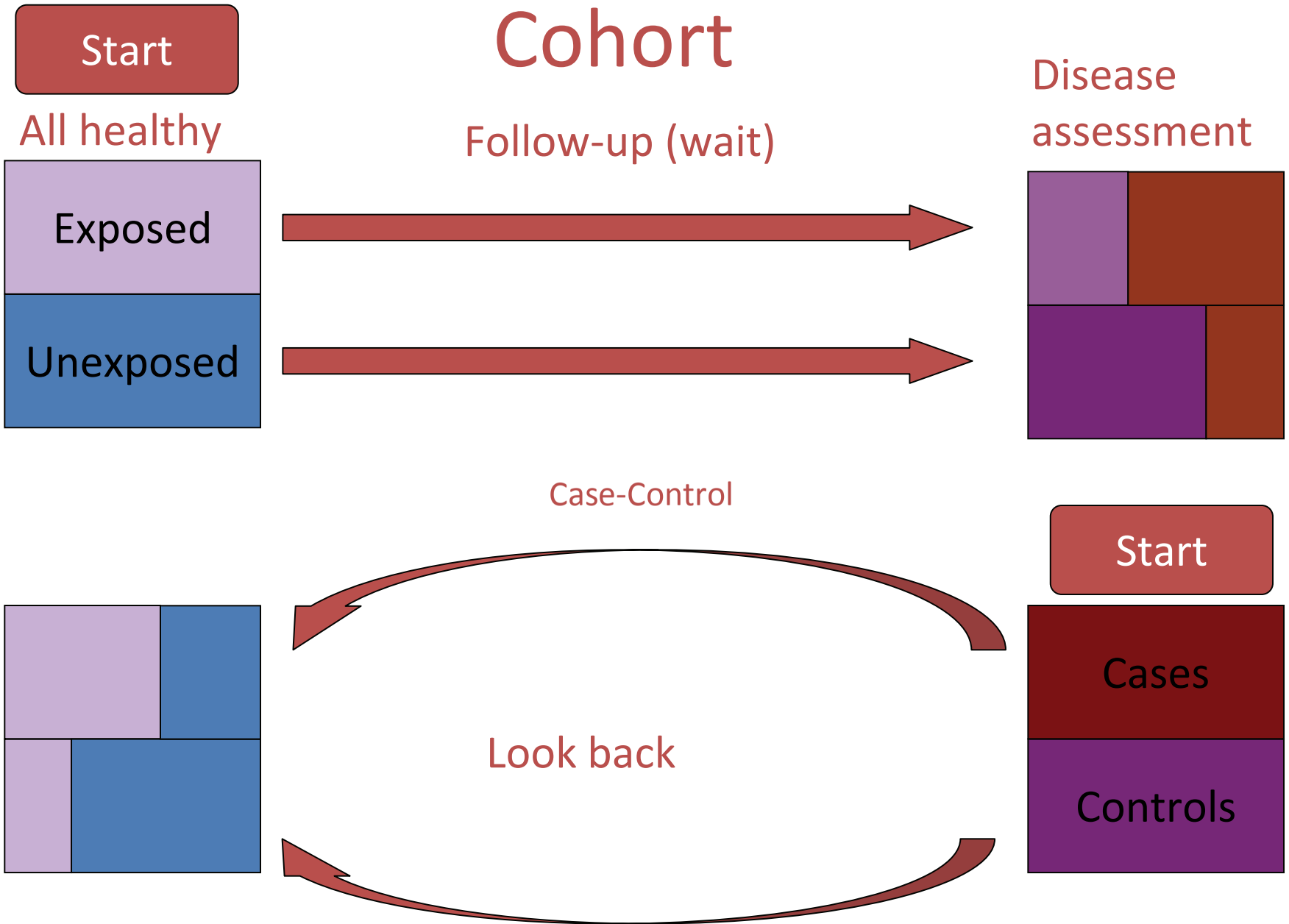
Selection Bias

- Selection of appropriate controls
  - hospital based studies
  - community based studies
  - Selection bias with cases
- Selection of cases
- Matched case-control studies
  - Cases and controls often differ in important aspects (age, sex, ethnicity, behaviours...)
  - These can confound the study
  - One way to eliminate such differences is matching controls to cases on these factors
  - More than 1 control per case can be used

Information Bias

- retrospective measurement of exposure may be problematic:
  - inaccurate - will underestimate any relation
  - reporting bias - overestimation
  - reverse causation - spurious association

Why carry out a case control study?


Cohort
Start
All healthy
Exposed
Unexposed
Follow-up (wait)
Disease assessment
Case-Control
Start
Cases
Controls
Look back

Strengths of case-control studies

- Quick (cases already exist, no need to wait)
- Cheap (not necessary to examine large number of people)
- Can examine many exposures
- Suitable to study rare diseases
- Suitable to study stable exposures (eg genetic markers)

Weaknesses of case-control studies

- Not suitable for rare exposure
- Cannot calculate incidence risk or death rates
- Prone to selection bias
- Prone to misclassification of exposure
- Prone to reverse causation (people with disease may change their behaviour)

References:

Doll R, Hill AB (1952) A study of the aetiology of carcinoma of the lung. *British Medical Journal* 2:1271–1286

Rothman K (2002) *Epidemiology. An Introduction*. Oxford University Press, Oxford, England

Schulz KF, Grimes DA (2002) Case-control studies: research in reverse;. *Lancet*: 359: 431–34