Epidemiological study design

Paul Pharoah

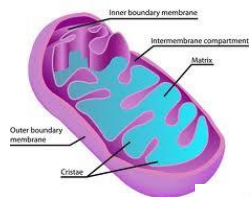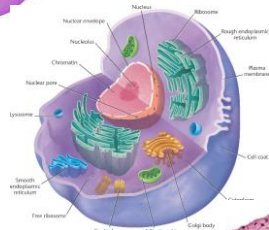Department of Public Health and Primary Care
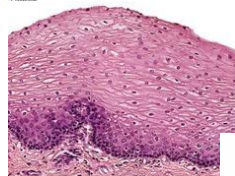
Caffeine

Molecules

What/why?

Mitochondrion

Organelles

Cells
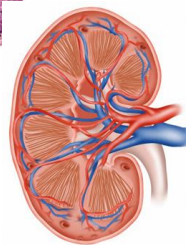
Tissues

Organs

Individuals

Clinical medicine

Public health medicine

Populations

Our knowledge of molecular biology, cellular biology, physiology and pathology is a fundamental part of clinical medicine.

In the end clinical medicine is practised on individuals and it is the cause of the disease in the whole individual and the response of the whole individual to treatment that matters.

Epidemiology is the study of the distribution, determinants and control of diseases in populations

In order to make inferences about individuals needs to sample many

Clinical epidemiology is the study of determinants of disease outcome in individuals with disease

Types of epidemiology

- Descriptive epidemiology
  - Study of distribution of health states
    - incidence, prevalence
  - Time
  - Place
  - Person

- Analytic epidemiology
  - study of the risk factors for health states

Analytic epidemiology

Exposure
Risk factor ⟶ Health
outcome

Is there an association?

What is the strength of the association?

Is the association causal?

Exposures

- Any factor that might be associated with outcome
- External environment
  - E.g. pollution, over-crowding
- Lifestyle factor
  - Diet, smoking
- Individual characteristic
  - Height, weight, blood pressure, blood cholesterol
- Medical intervention
  - Drug, surgery

Outcomes

- Any health related state
- Disease occurrence in healthy person
- Health outcome in person with disease
  - Disease complications, survival etc
- Occurrence of a physiological trait
- An exposure can also be an outcome
  - Smoking behaviour, BMI

Always be clear when thinking about study design what the main exposure of interest is and what the main outcome of interest is

Key terms

- Population
- Prospective and retrospective
- Longitudinal
- Observational and experimental
- Ecological, cohort and case-control studies
- Relative risk and odds ratio

Study designs

Exposure and outcome measured at population level

Exposure and outcome measured in individuals

Exposure as occurred in free-living people

Exposure assigned experimentally

Observational study
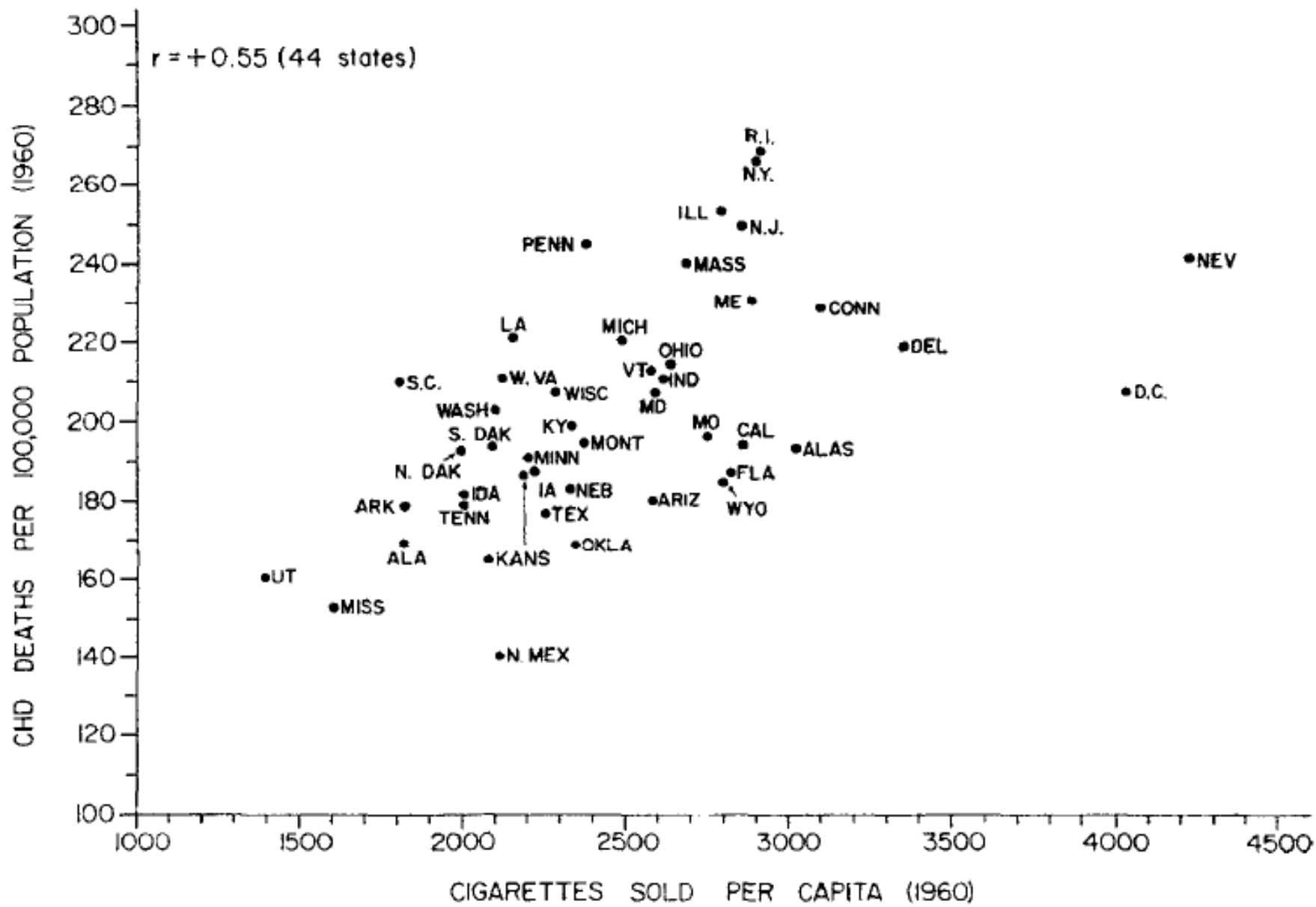
Ecological study

Case-control study

Cohort study

Clinical trial

Population

- Defined group of people
  - E.g. Everybody in Brazil
  - All women aged 50 to 69 in Cambridge
- Important to be clear what the relevant population is
- What is the study population
  - generalisability of results
  - e.g. trials of cholesterol lowering agents in men and applicability of results in women
- What is the population of interest

The ecological study

- Summary measures of exposure and outcome obtained for different populations
- Test for correlation of exposure and outcome at population level
- Observational study

GD Friedman. *J Chronic Dis*, 1967.
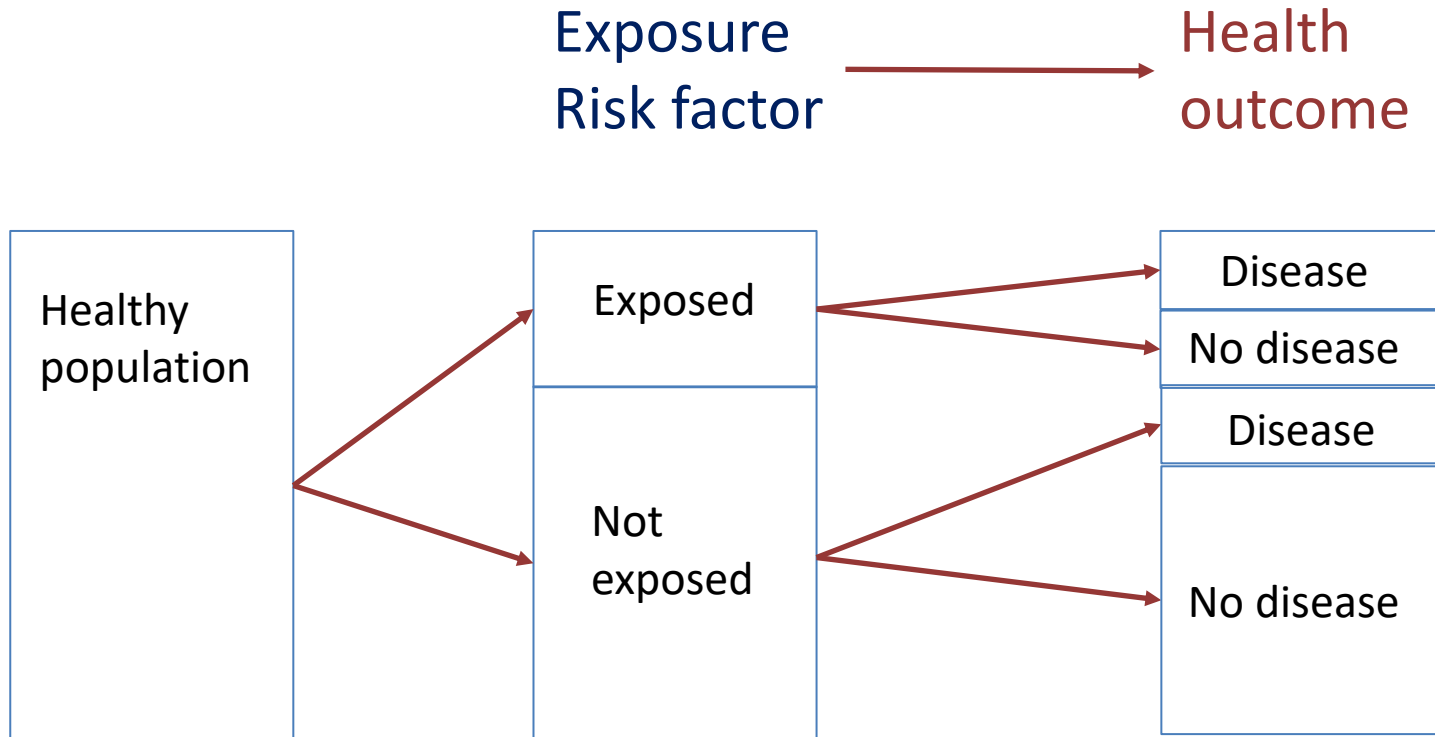
Pros and cons of ecological study

Advantages

- Easy to do
- Based on routine data
- Good for hypothesis generation

Disadvantages

- Relies on available exposure and outcome measures
- Only single exposure
- Confounding a major problem

Cohort study

Exposure
Risk factor $\longrightarrow$ Health outcome

Population defined by observed exposure status at start of follow-up
  – observational study
Population followed-up over time to observed outcome status
  - longitudinal study
  - prospective study
  - cohort study

The retrospective cohort

- Population of interest defined after follow-up time has already occurred
  - E.g. Ovarian cancer patients diagnosed 2000-04
- Exposure status at time of entry into follow-up determined
  - E.g. Treatment
- Outcome status at end of follow-up determined (has already happened)
- A cohort or longitudinal study
- Retrospective refers to the ascertainment of exposure status
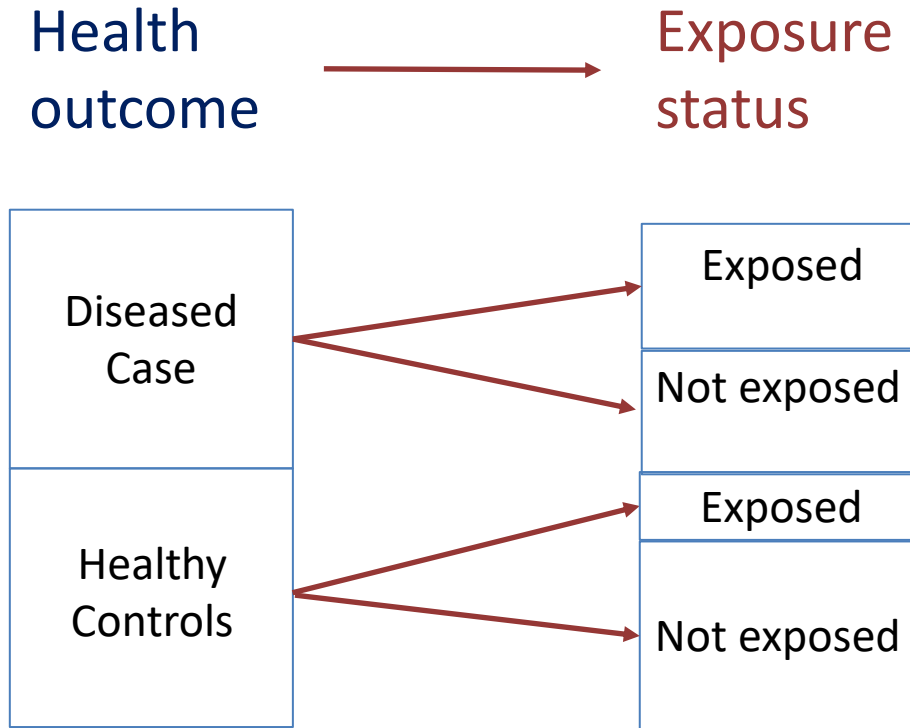
Pros and cons of cohort study

Advantages

- No bias in exposure
- Exposure precedes outcome
- Can investigate multiple exposures
- Can investigate multiple outcomes
- Direct estimate of incidence
- Good for rare exposures

Disadvantages

- Time to do study
- Not good for rare outcomes
- Large sample size
- Loss to follow-up
- Potential bias in outcome

Case-control study

Health outcome → Exposure status

		Exposed
Diseased Case	→	Not exposed
Healthy Controls	→	Exposed
		Not exposed

Study population defined by observed outcome status
- take samples of cases and sample of controls
– observational study
Exposure status then determined
- retrospective study

Pros and cons of case-control study

Advantages

- No bias in outcome
- Can investigate multiple exposures
- Good for rare outcomes
- Rapid (no follow-up)
- Efficient / less costly

Disadvantages

- Prone to bias in exposure measurement
- Not good for rare exposures
- Time relationship between O & E unclear

The clinical trial

- Special example of a prospective cohort in which exposure status is assigned to individual
- Experimental study
- Often very limited eligibility criteria
  - the trial "population" may not be representative of the patient "population" of interest
- Random allocation of exposure (intervention) reduces probability of confounding
- Blinding of participant and investigator prevents information bias

Pros and cons of clinical trial

Advantages

- No bias in exposure
- No selection bias
- Blinding can minimise bias bias in outcome ascertainment

Disadvantages

- Single exposure
- Not good for rare outcomes
- Generalisability
- Randomisation may be difficult/unethical
- Cost
- Follow-up time

The 2 x 2 table

		Outcome status		
		Disease	No disease	Total
Exposure status	Exposed	a	b	a + b
	Not exposed	c	d	c + d
	Total	a + c	b + d	a + b + c + d

Measure of association

- Statistical test – is association present or not
- Estimate a parameter to provide a measure of strength of association
  - relative risk (a.k.a. risk ratio) from cohort study
  - odds ratio from case-control study
- Odds ratio from a case-control study is approximately equal to the relative risk if the disease is uncommon in population
  - rare disease assumption

Relative risk and odds ratio

		Outcome status		
		Disease	No disease	Total
Exposure status	Exposed	a	b	a + b
	Not exposed	c	d	c + d
	Total	a + c	b + d	a + b + c + d

$$\text{Relative risk} = \frac{a / (a + b)}{c / (c + d)}$$

$$\text{Odds ratio} = \frac{a / c}{b / d} = \frac{a.d}{b.c}$$

Explanations for observed association

1. Chance – a statistical fluke
   - Possible explanation for every association
2. Bias
3. Confounding
4. True association

Bias

- Systematic deviation from truth producing a mistaken estimate of the relationship of an exposure with the risk of disease
- Two main types of bias
  - selection bias
    - occurs when participant selection that distorts the exposure-outcome relationship from that present in the target population
  - information bias
    - occurs when information is collected differently between two groups, leading to an error in the conclusion of the association

Examples of selection bias

- Observed effect of disease screening could be due to those attending screening being more health conscious than non-attenders
- Study health of workers in a workplace exposed to some occupational exposures comparing to health of general population

  Working individuals are likely to be healthier than general population that includes unemployed people (Healthy worker effect)
- Healthy migrant effect – comparing migrants groups to non-migrant groups
- Using hospital based controls in a case-control study

Examples of information bias

- Recall bias in case-control studies when cases may be more likely to recall an exposure than controls
- Determination of outcome might be influenced by exposure status
 - e.g. more intensive follow-up of group of individuals with particular exposure

Bias and study design

- Bias can create a spurious association or hide a true association
- Careful study design can minimise bias
- Retrospective studies more prone to bias
  - just because a study has a retrospective design does not mean there is bias
- Even clinical trials can have information bias
  - minimised by blinding of participants and investigators to randomly allocated intervention

Confounding

Exposure
Risk factor
- - - - - - - - - → Health
outcome
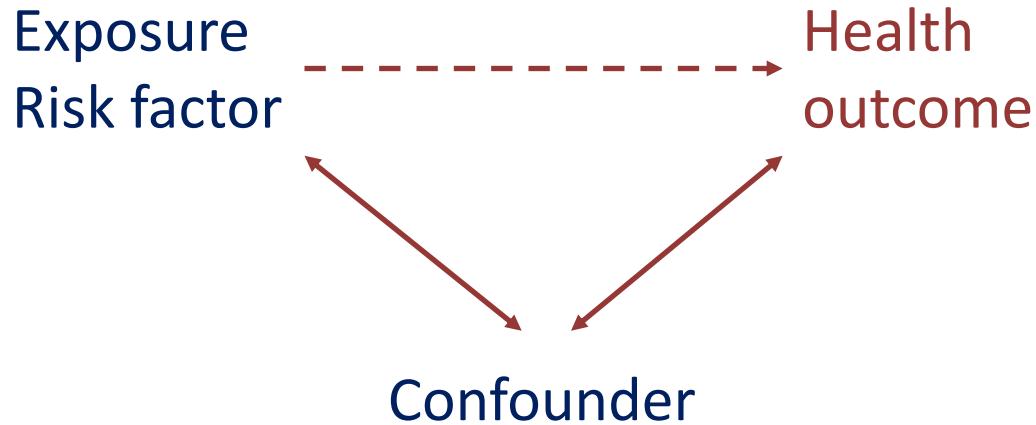
Confounder

A confounder is a factor that is associated with exposure AND outcome of interest

Coffee
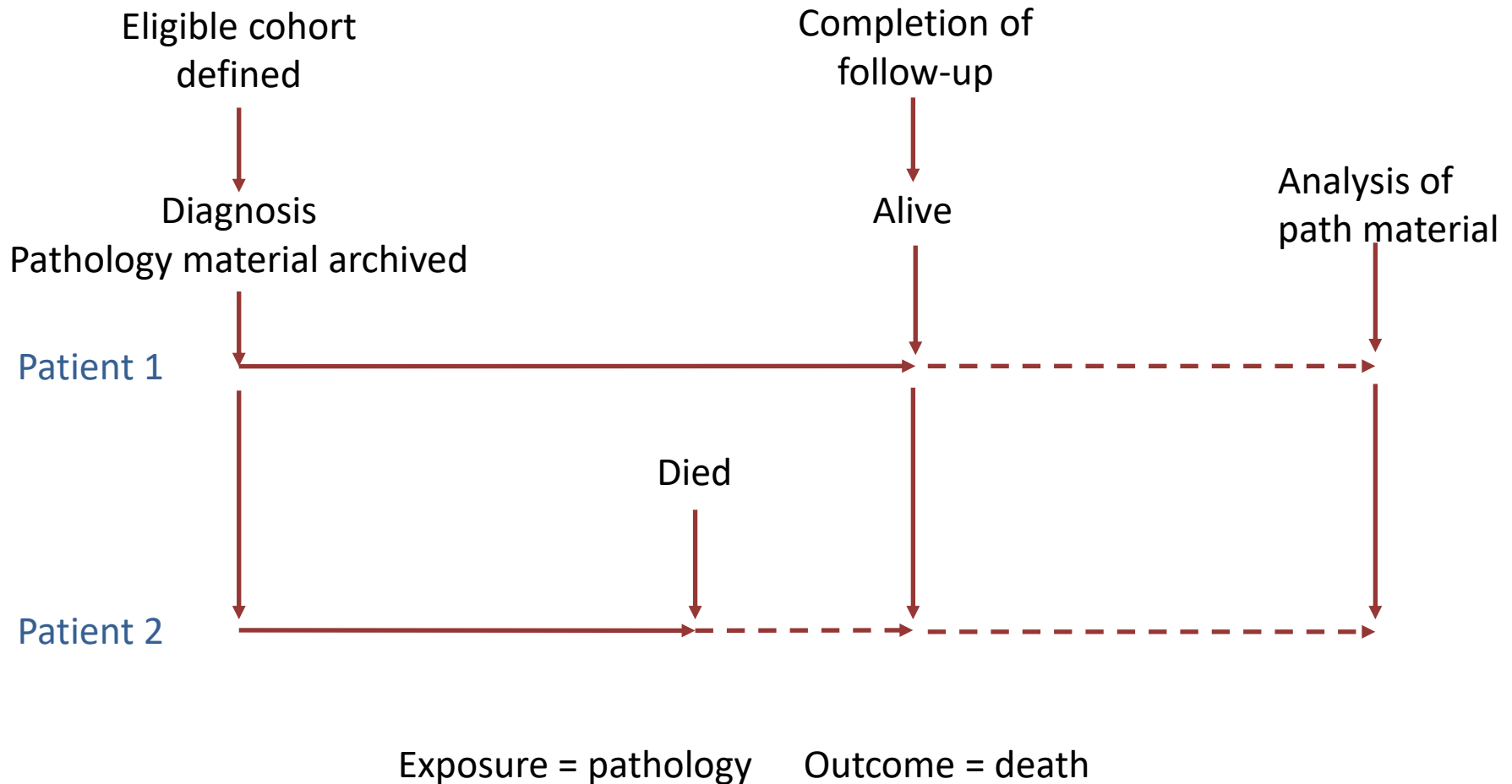drinking
- - - - - - - - - → Lung
cancer

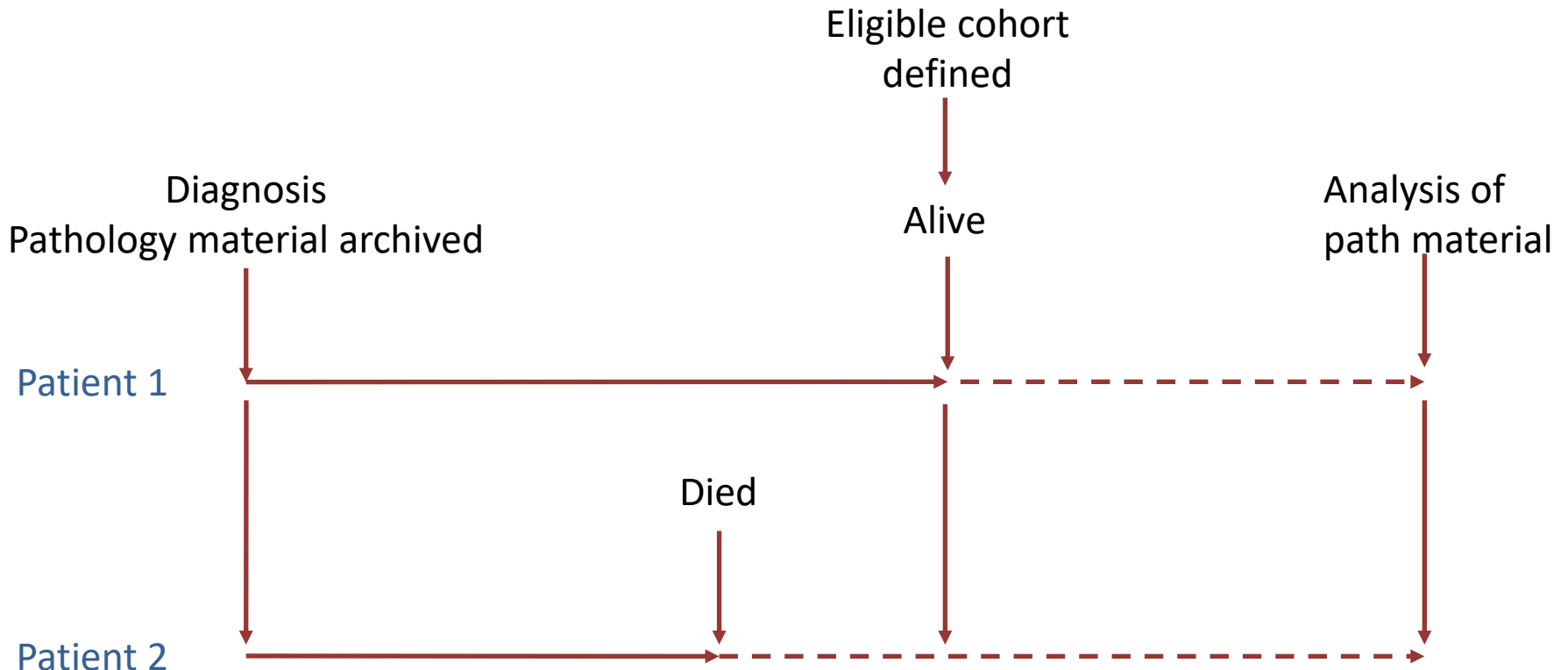Smoking
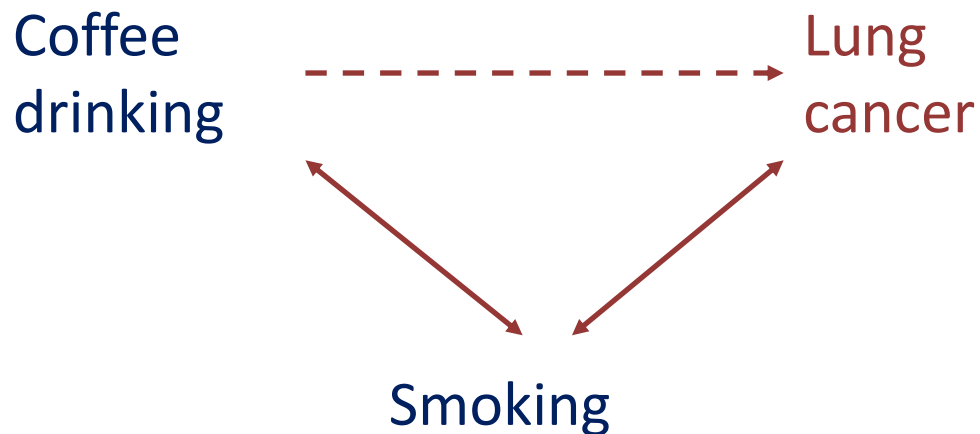
Confounding and study design

- Many possible causal exposures are correlated
  - age
  - sex
- Major problem in observational studies
- Can match for confounding in study design
- Can control for confounding in analysis
- Cannot manage a confounder that you do not know about or cannot measure
- Primary purpose of a randomised clinical trial is to remove confounding
  - all potential confounding factors are equalised between two groups

An example of clinical cohort study: "prospective" cancer case series

An example of clinical cohort study "Retrospective" cancer case series

Eligible cohort
defined

Diagnosis
Pathology material archived

Alive

Analysis of
path material

Patient 1

Died

Patient 2

Some blurring of the distinction between pro- and retrospective.
Biases of most retrospective studies not relevant

Issues

- No study design is perfect
- Do not use prospective and retrospective as synonyms for cohort and case-control
- Possible biases may or may not be present
- Some epidemiological principles apply to laboratory study design

Questions or comments?