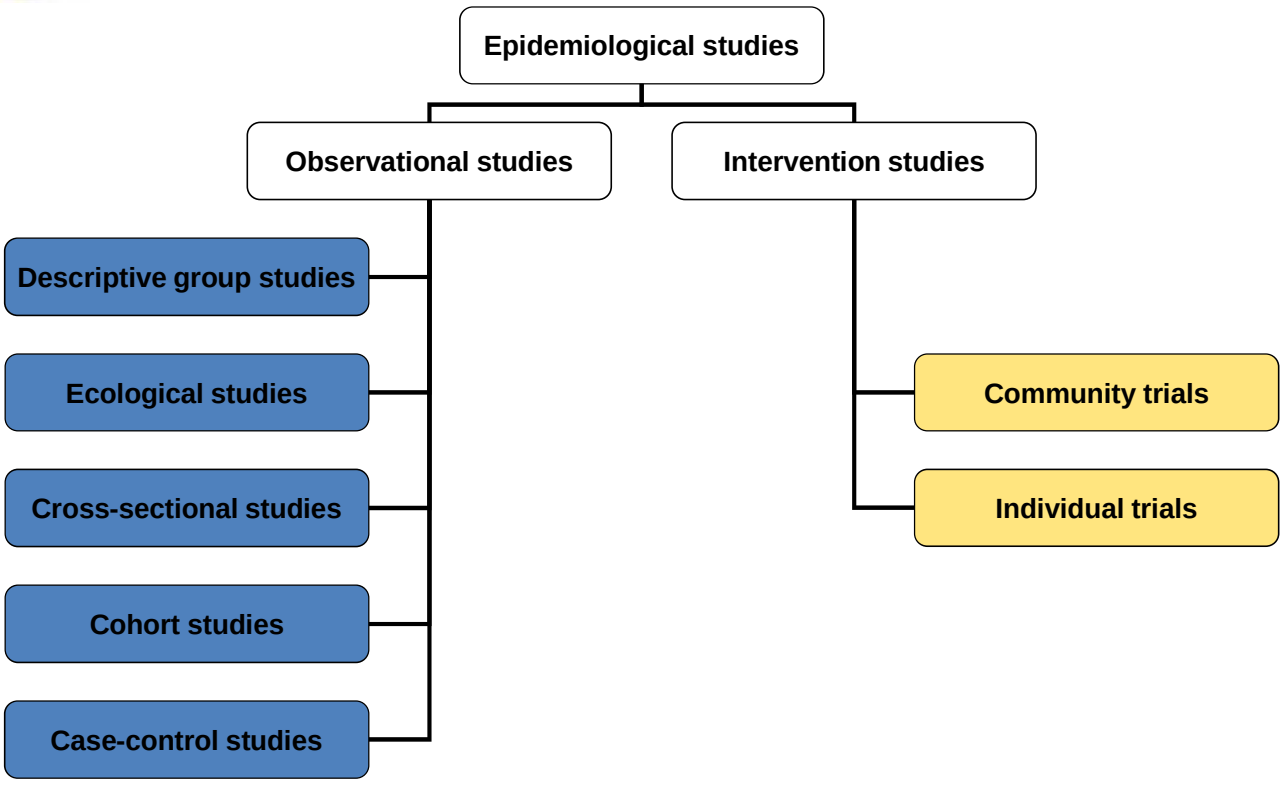
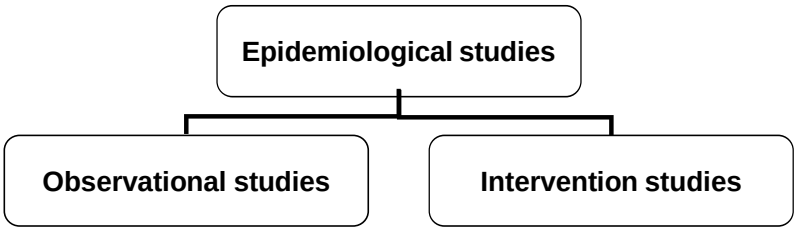
Lecture 2

Study designs
for research on SDH

Overview



Observational vs intervention



We observe only "non-experimental"

We allocate exposure "experimental"

Key definitions

1. Outcome of interest
   - The event or state for which we want to know more about its causes
2. Exposure of interest
   - The "risk factor" under study hypothesised to influence the outcome of interest
3. Other exposures that may influence the outcome (potential confounders)

Confounding

Alcohol during pregancy

Exposure of interest

Low birth weight baby

Outcome

Smoking during pregancy

Confounder

Exposures and outcomes

An **outcome** in one study, can be **exposure** in another

Low birth weight

Exposure of interest

High blood pressure

Outcome

Descriptive group studies

- Describing the frequency of certain factors (e.g. disease, mortality, etc.) in populations
- Person – place – time
- Often from routine data collection
- E.g. mortality statistics

Ecological studies

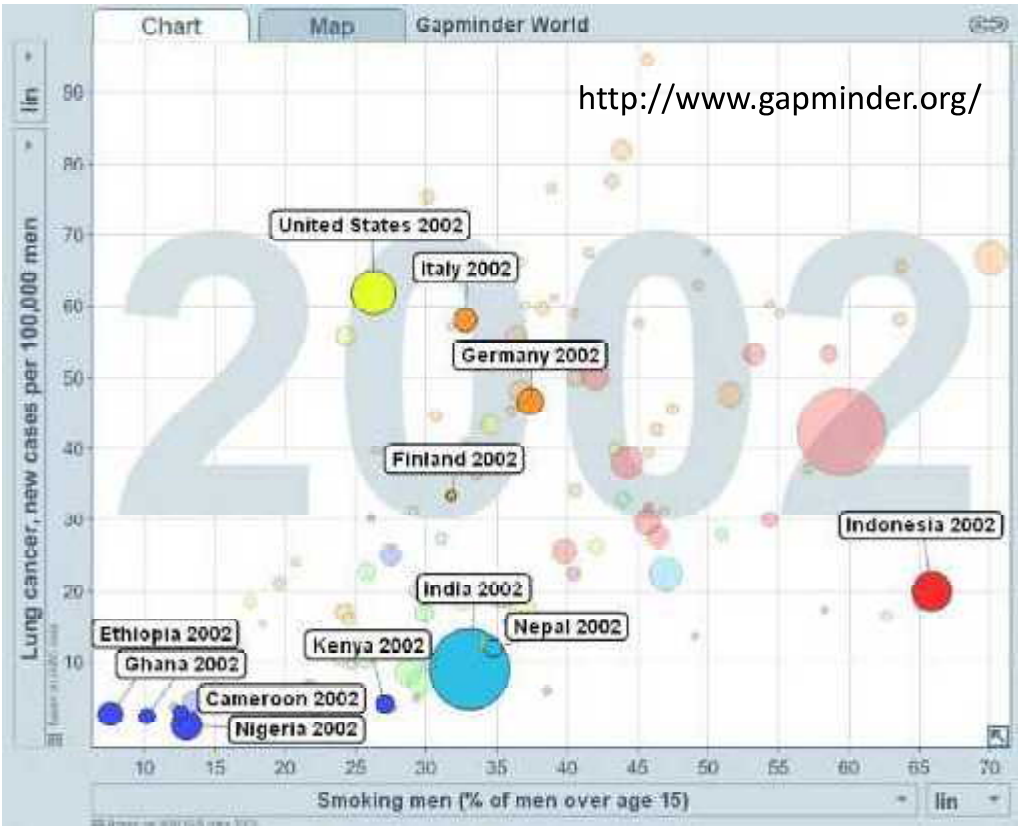
= "Studies in which the units of analysis are populations or groups of people, rather than individuals."

Last: Dictionary of Epidemiology, 1988

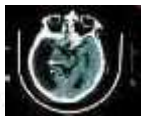
Ecological studies

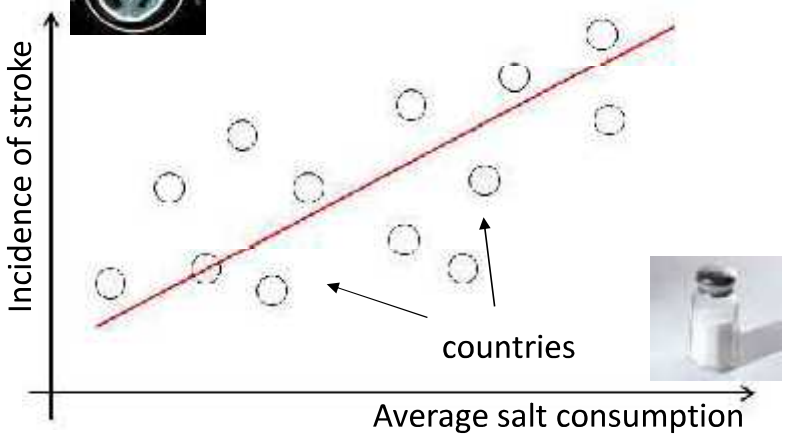
- Describing disease frequency in populations
- But additionally including information on an exposure
- Comparison of average disease frequency between groups with different average exposure
- Looking at association → = analytical study
- Has nothing to do with "ecology"
- Better term may be "correlation studies"

Ecological studies – example 1



Ecological studies – example 2





www.who.int
Cardiovascular diseases (CVDs) are the leading cause of death in the world. High blood pressure is the leading risk for mortality globally.

The WHO Member States in WHA 66.10 have agreed on a voluntary global NCD target for a 30% relative reduction in mean population intake of salt, with the aim of achieving a target of less than 5 grams per day (approximately 2g sodium) by 2025.
They have also agreed on a voluntary global NCD target for a 25% relative reduction in the prevalence of raised blood pressure.

Chocolate and nobel prizes

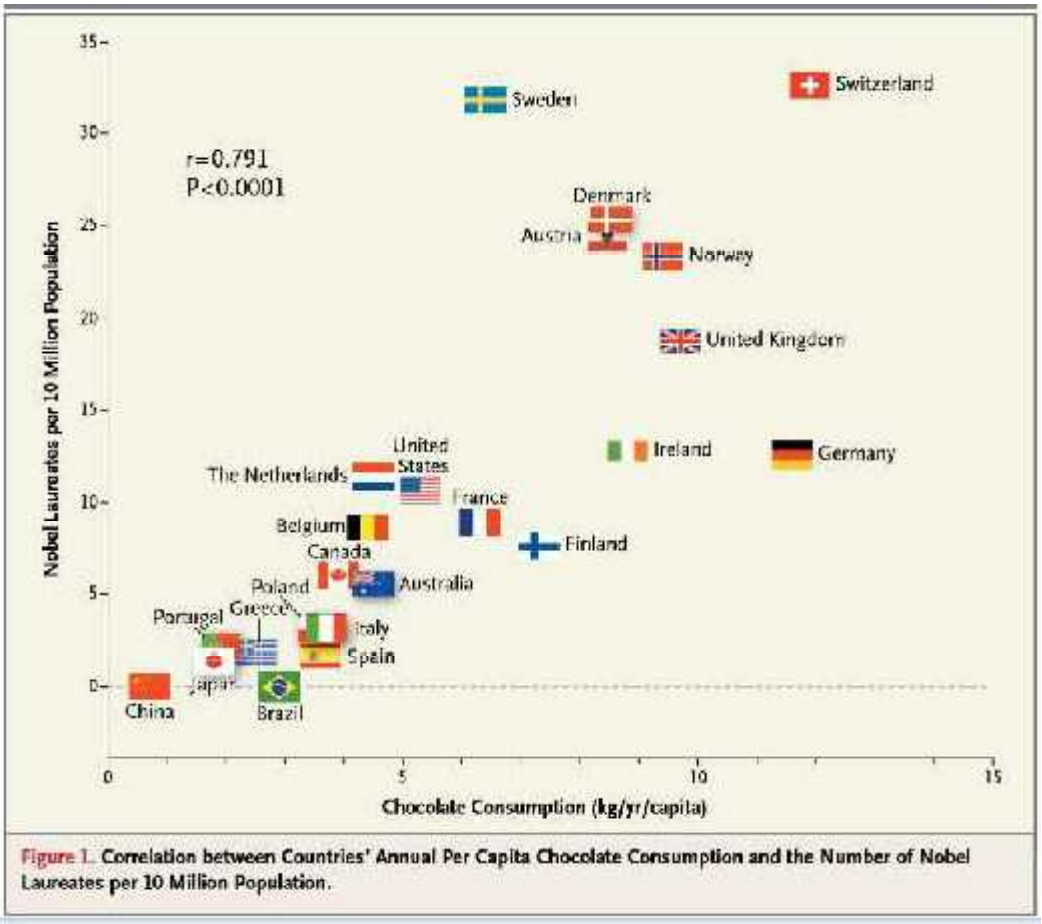
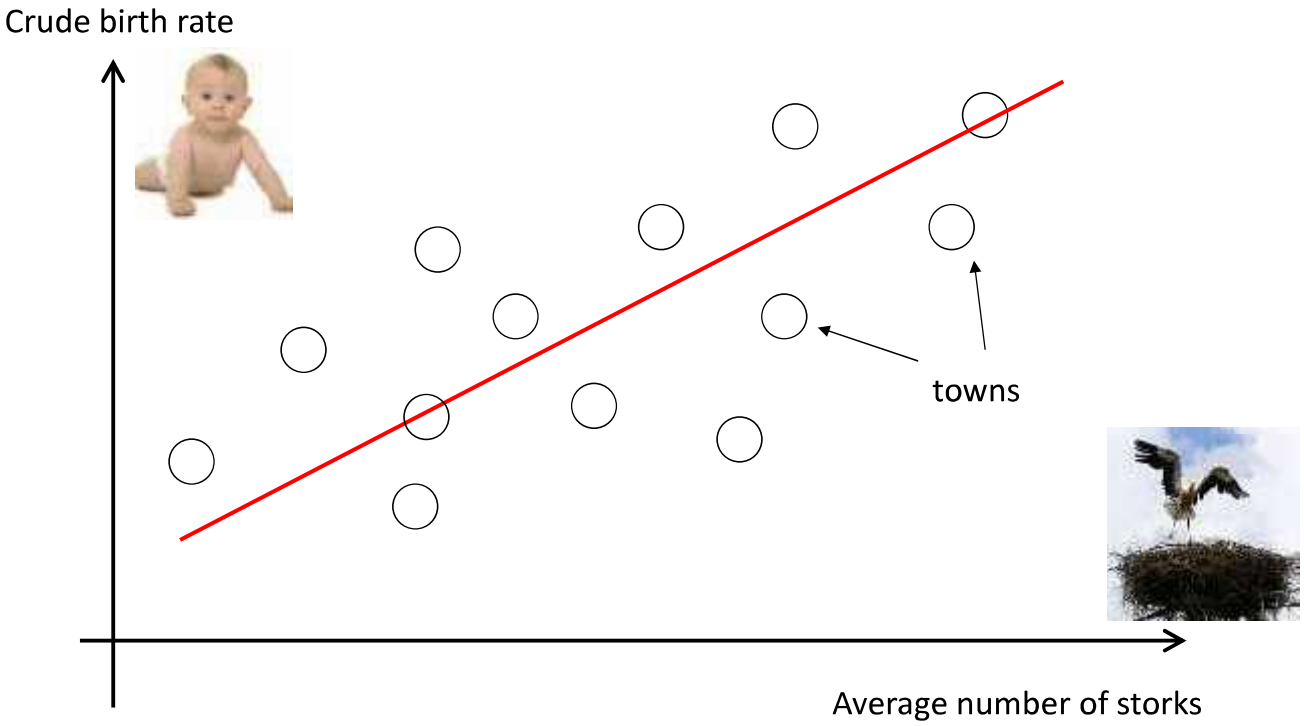


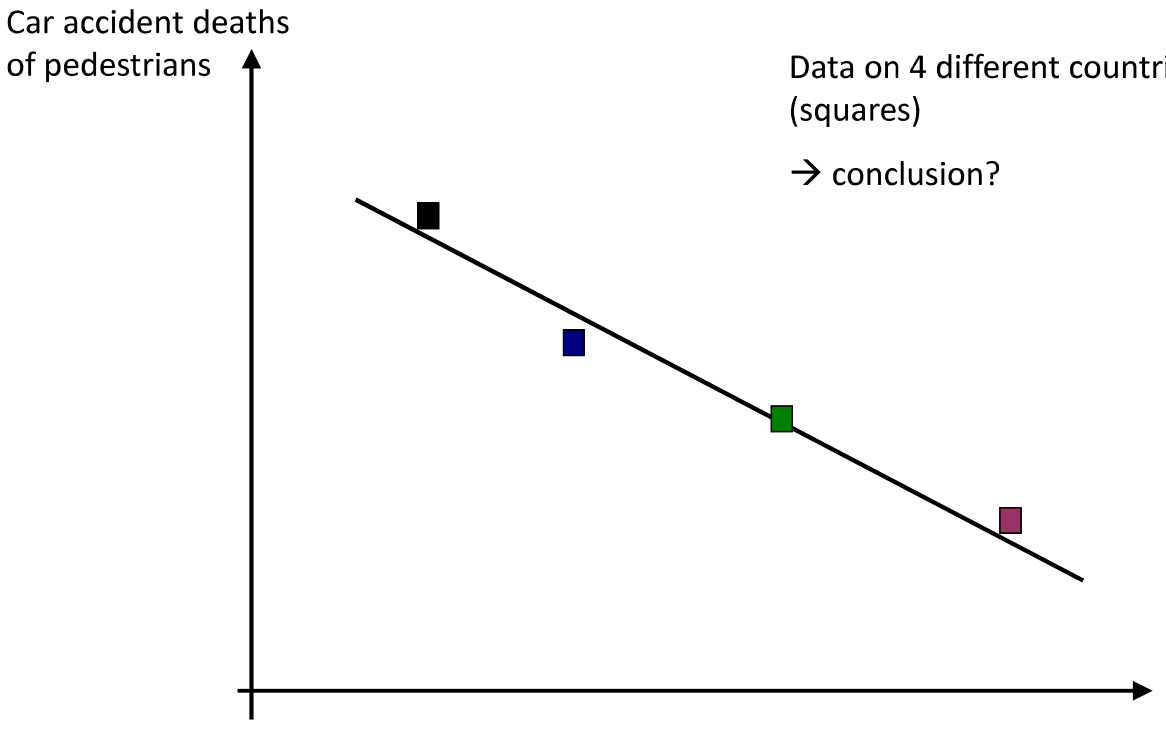
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Messerli, NEJM (2012) 367;16

Ecological studies – example 3

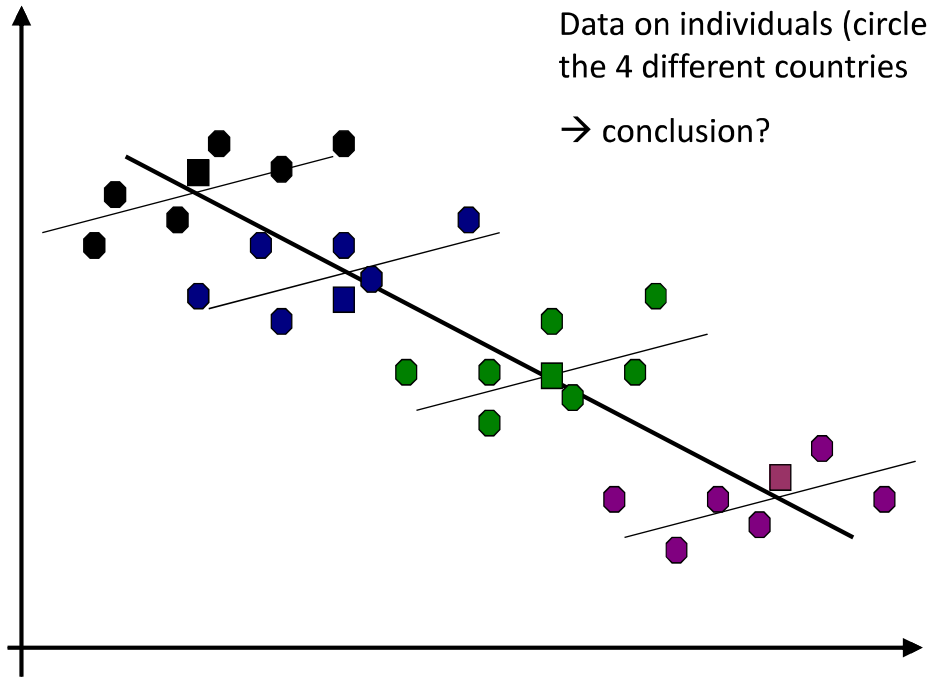


Ecological studies – example 4



Ecological studies – example 4

Car accident deaths of pedestrians



Data on individuals (circles) in the 4 different countries

→ conclusion?

Poverty

Ecological studies

+ Quick and cheap
+ Useful for studying group-level properties e.g. health care system
+ Useful for hypothesis generation (especially when difference between groups is larger than within groups)
- Limited control for confounding
- Group results cannot be extrapolated on the individual level ("ecological fallacy")

  Historical example: Durkheim 1897 in Prussia: higher suicide levels in mainly protestant regions

→ Useful but one needs to be aware of the limitations

Cross-sectional studies

- At one **point** in time → Does **NOT** measure incidence!
- **Descriptive**: prevalence of a variable of interest
- **Analytical**: prevalence of outcome and exposure
  → comparing prevalence of the outcome between exposure groups: association?

Exposure and outcome are measured at the **same** point in time!

→ This creates some problems…

Cross-sectional studies

Exposure and outcome are measured at the **same** point in time

→ Is the exposure we measure now really the aetiologically relevant exposure?

Example: Risk factors for cervical cancer?

1. blood group
2. working with chemicals
3. sexual activity

Latency period exposure→outcome

Change over time

→ Ask about past exposure (but: recall errors,… more later)

Cross-sectional studies

And furthermore…

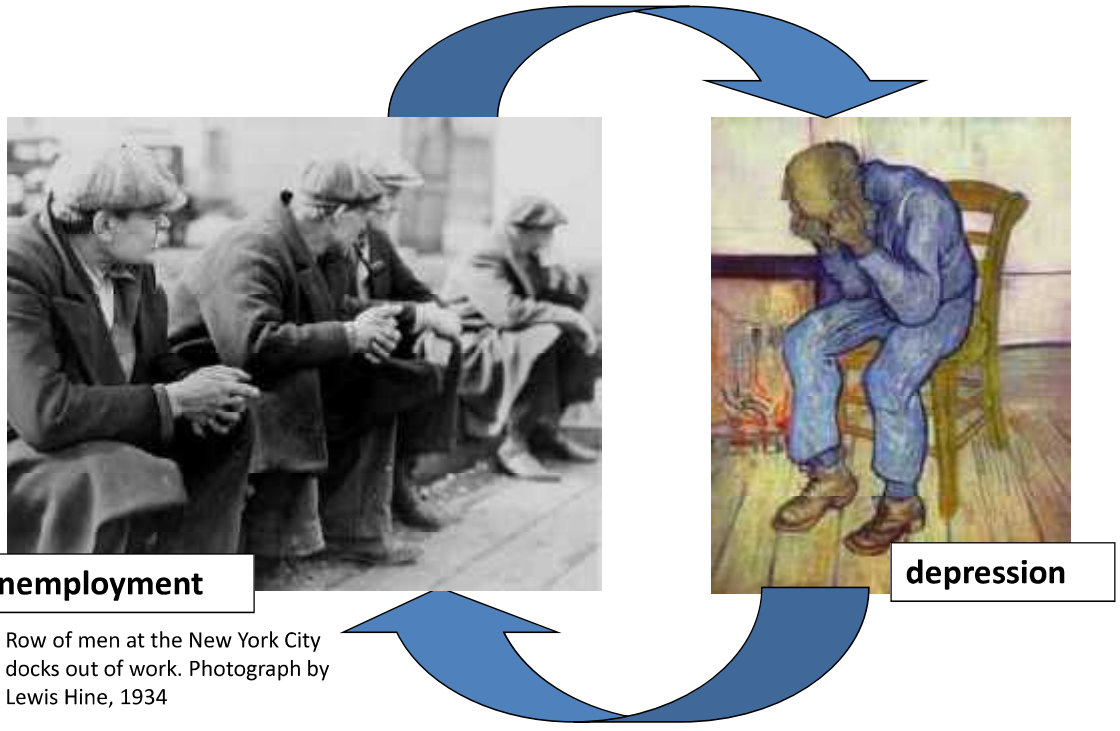
→ Do we really know the exposure came first or did having the outcome change the exposure?

→ "chicken and egg" problem

<u>Example:</u> Does unemployment cause depression?

Maybe, but depression may also make people more likely to lose their job…

Reverse causality



Row of men at the New York City docks out of work. Photograph by Lewis Hine, 1934

Reverse causality

Methodological issues

- **Validity** of the measurements: Do we really measure what we want to measure?
- **Repeatability** of the measurements: Do we get the same results if we measure the same?
- **Sampling**: Representativeness of the sample
  → random sample preferred

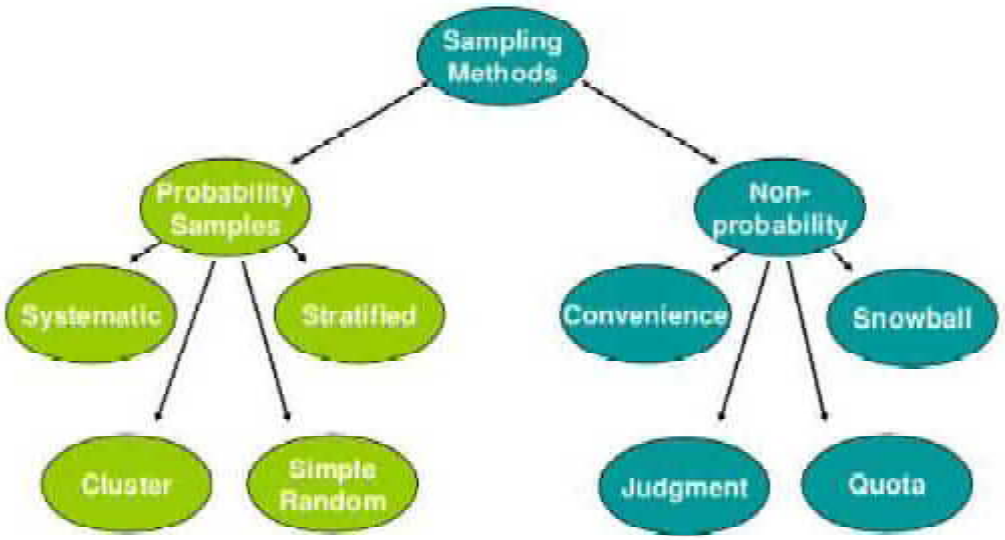
Example: What is the prevalence of anaemia in women in Accra?

- Sampling among women encountered at the exit of the gym?
- Sampling among women encountered in a supermarket?
- Sampling among women encountered at the university campus?
- Sampling among women who have a landline telephone number?

Sampling

- Steps
  - Definition of target population
  - Selection of a sampling frame (list)
  - Probability or Nonprobability sampling
  - Sampling Unit
  - Error
- Random sampling error (chance fluctuations)
- Non-sampling error (design errors)

Classification of sampling methods



Methodological issues

- **Non-response**:

  Even if we picked a random sample, what happens if 25% of people refuse to join the study?

Example: Of 1000 women in our anaemia study,

250 refused to provide a blood sample = response rate 75%

- We find 75 / 750 = 10% to be anaemic.

- If all the non-responders were anaemic, what would we have found?

  75+250 / 1000 = 32.5% anaemia prevalence

- If all the non-responders were non-anaemic, what would we have found?

  75 / 1000 = 7.5% anaemia prevalence

Methodological issues

- Measurement problems (validity and repeatability of our instruments)

  → "information bias"

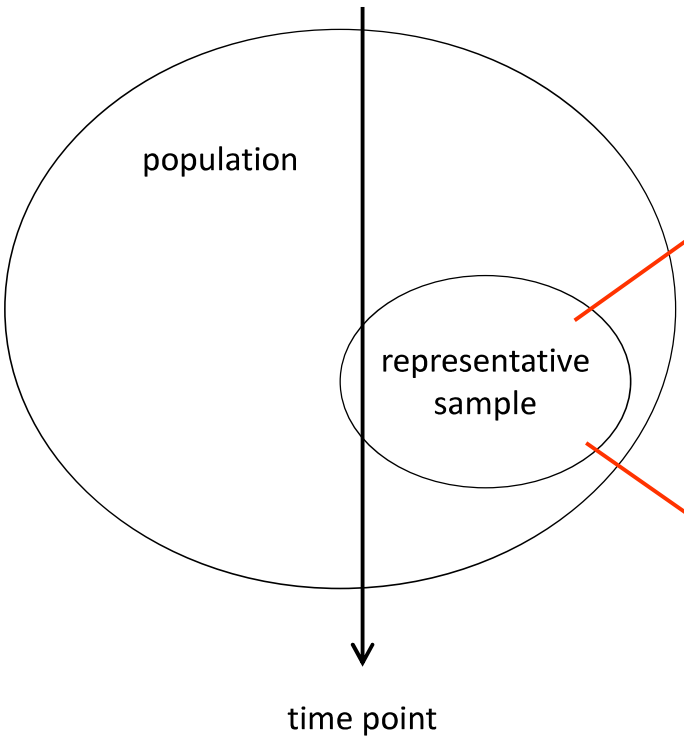
- Sampling and non-response problems

  → "selection bias"

BIAS = systematic error → wrong estimate

(more later)

Cross-sectional studies

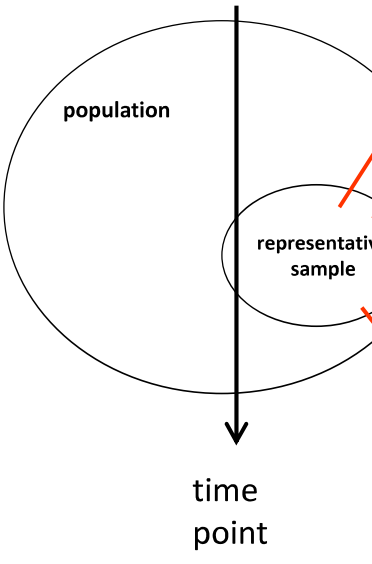


Number **WITH** the factor of interest (disease, risk factor,...)

Calculate **prevalence** (measure of **frequency**)

Number **WITHOUT** the factor of interest (disease, risk factor,...)

Cross-sectional studies



Number **WITH** outcome and **WITH** exposure

Number **WITHOUT** outcome but **WITH** exposure

Calculate **prevalence** among the exposed

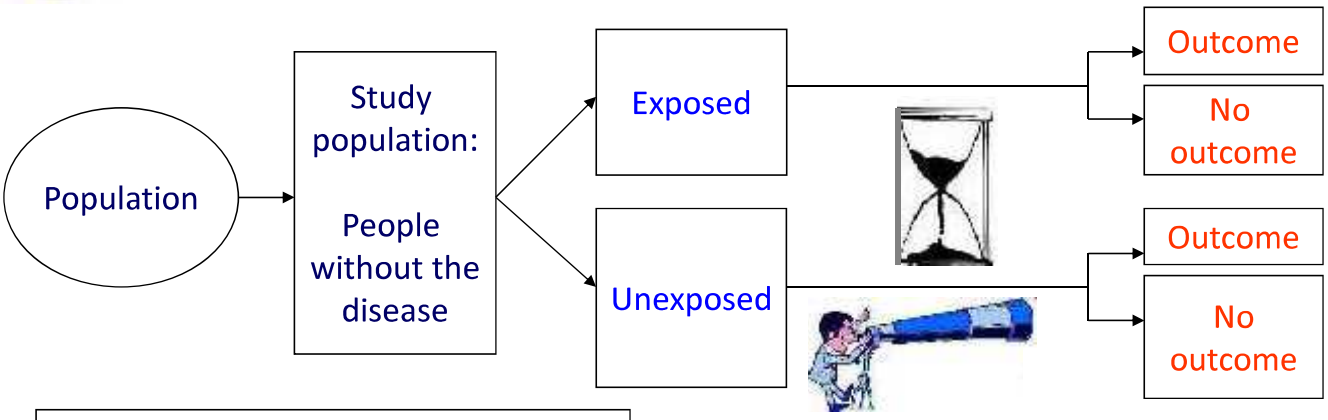
Calculate **prevalence ratio** or **difference** or **odds ratio** (measure of **effect**)

Number **WITH** outcome but **WITHOUT** exposure

Number **WITHOUT** outcome and **WITHOUT** exposure

Calculate **prevalence** among the unexposed

Cohort study design



Proceeding in cohort studies:

→ Classification of individuals as exposed / unexposed

→ Following up the entire study population over time

→ Comparison of incidence of the outcome in the exposed and unexposed individuals

Direction of enquiry

Time

Cohort studies

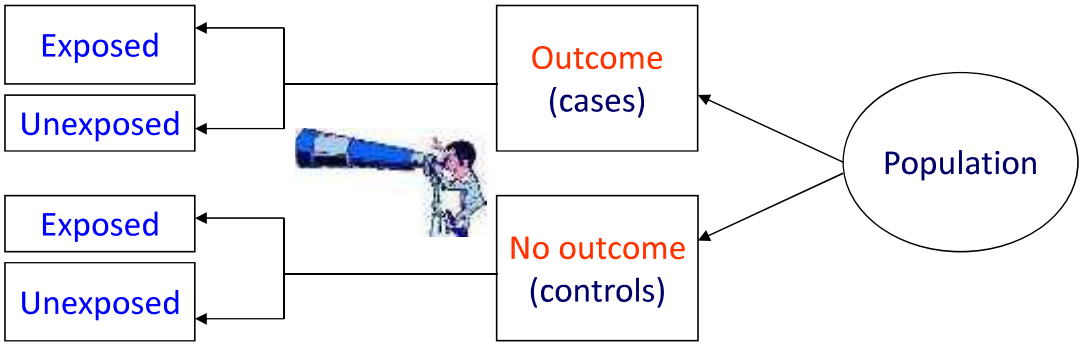
= Follow-up studies

- Follow individuals who are initally free from the outcome (e.g. a certain disease) over time
- Measures **incidence** of an outcome (or several outcomes)
- Exposure is determined BEFORE development of the outcome

→Overall incidence (descriptive study)

→Comparing incidence between exposed and unexposed (analytical study)

Case-control study design



Direction of enquiry

Time

Proceeding in case-control studies:

→ Selection of subjects on the basis of their outcome status

→ Finding out the exposure status of both cases and controls

→ Comparison of exposure status between cases and controls

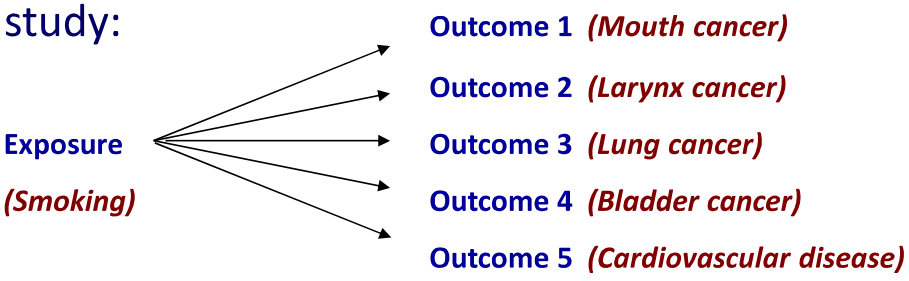
Case-control studies

- "Backwards": Asking about the exposure once the outcome has already developed
- Case selection: purposefully picking people with the outcome of interest
- Control selection: from the population that gave rise to the cases (tricky!)
- Useful for rare outcomes ("concentrating" cases)
- No information on measures of frequency (unless we know the sampling fraction for cases and controls)

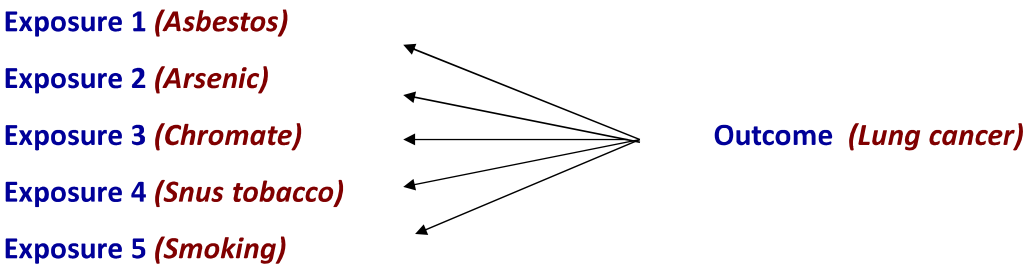
  → always analytical
- Can investigate several different exposures

Comparison cohort and case-control study

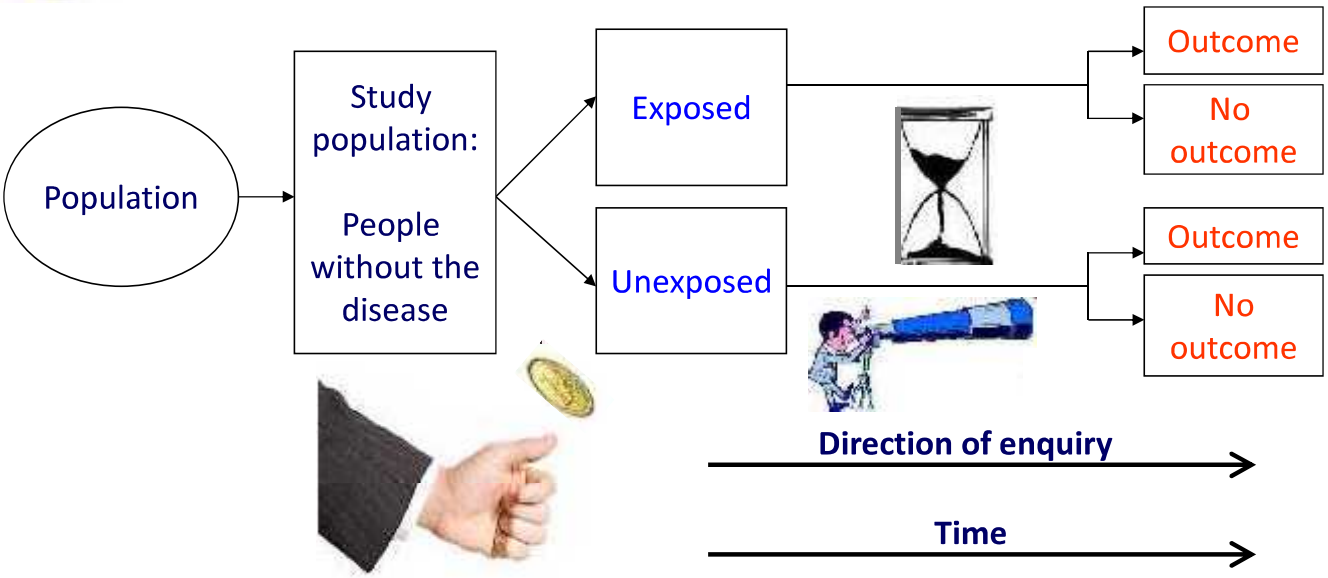
Cohort study:



Case-control study:



Intervention study design



Proceeding in intervention studies:

→ **Allocation** of individuals or groups into exposed / unexposed

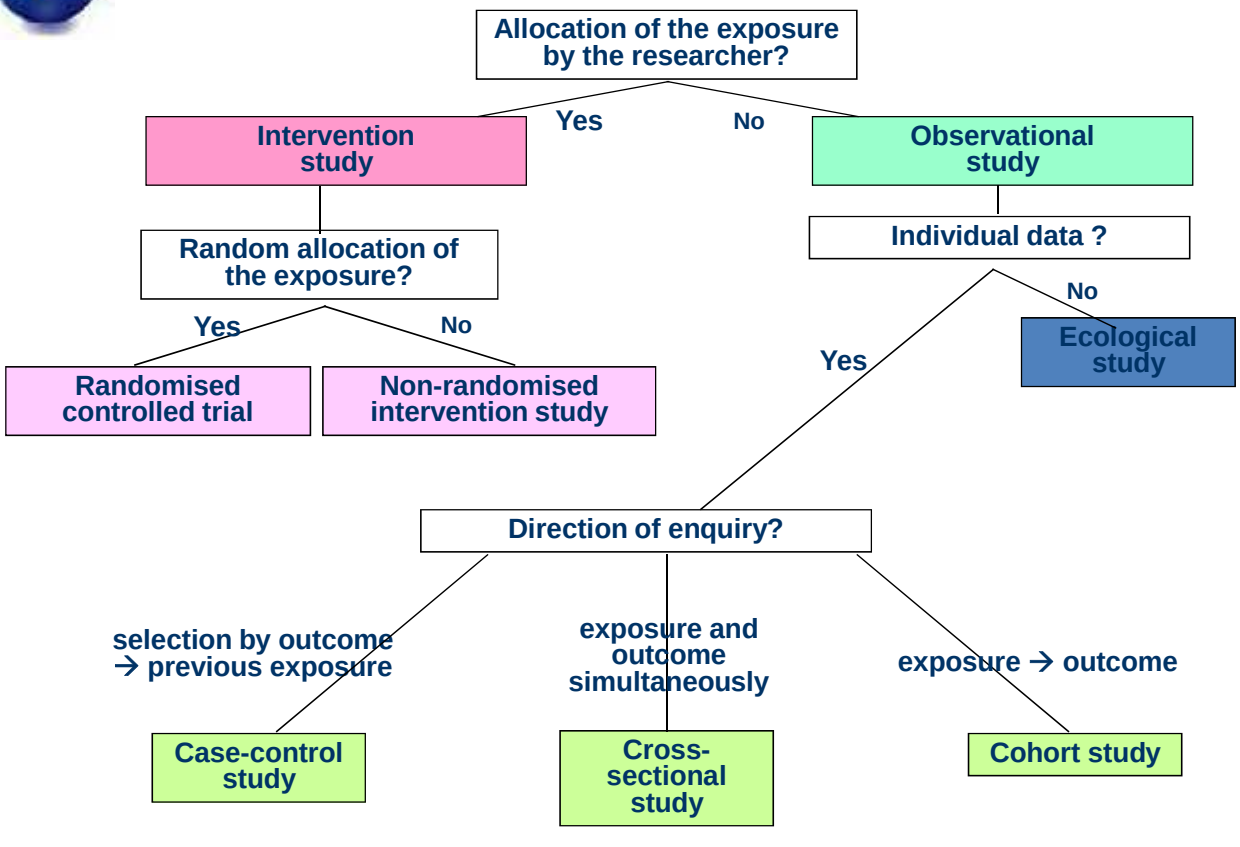
→ Otherwise similar to cohort

Intervention study

- The researcher deliberately allocates the exposure
- Not always possible for ethical reasons
- Preferred form: randomised controlled trial (RCT)
- Allocation by chance alone
- Individual allocation (clinical trial) or group allocation (community trial)
- If the trial is sufficiently large, all other risk factors will be evenly distributed between the groups → no confounding
- Set up to study effect of exposures in a clean setting

  → always analytical

Algorithm



Choice of study design

Which is the best study design?

→ This depends on the question and circumstances!

On what does it depend exactly?

- Frequency of the outcome, frequency of the exposure
- Whether want to study several outcomes, or several exposures
- Whether reverse causality is an issue
- Whether we need an incidence measure
- Expected latent period between exposure and outcome
- Time scale and funding for study
- Issues of bias and confounding
- What studies have been done before already, what is known