

# Choosing the right study design

**Caroline Sabin**

**Professor of Medical Statistics and Epidemiology  
Institute for Global Health**

# Conflicts of interest

---

I have received funding for the membership of Data Safety and Monitoring Boards, Advisory Boards and for the preparation of educational materials from:

- Gilead Sciences
- ViiV Healthcare
- Janssen-Cilag

# Main types of study design

---

Randomised controlled trial (RCT)

Cohort study

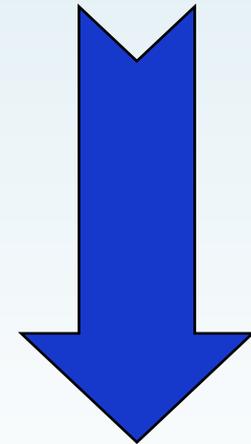
Case-control study

Cross-sectional study

Case series/case note review

‘Expert’ opinion

**BEST QUALITY  
EVIDENCE**



**WORST QUALITY  
EVIDENCE**

# Experimental vs. Observational

---

## *Experimental study*

Investigator **intervenes** in the care of the patient in a **pre-planned, experimental way** and records the outcome

## *Observational study*

Investigator does not intervene in the care of a patient in any way, other than what is routine clinical care; investigator simply **records** what happens

# Cross-sectional vs. Longitudinal

---

## *Cross-sectional study*

Patients are studied at a **single time-point only** (e.g. patients are surveyed on a single day, patients are interviewed at the start of therapy)

## *Longitudinal study*

Patients are followed over a **period of time** (days, months, years...)

# Assessing causality (Bradford Hill criteria)

---

- Cause should precede effect
- Association should be plausible (i.e. biologically sensible)
- Results from different studies should be consistent
- Association should be strong
- Should be a dose-response relationship between the cause and effect
- Removal of cause should reduce risk of the effect

# Incidence vs. prevalence

---

**Incidence:** proportion of patients **without the event of interest** who **develop the event** over the study period

- Can only estimate from a longitudinal study
- Must exclude those who have the event at start of study from the calculation

**Prevalence:** proportion of **all patients in study** who have the event **at a particular point in time**

- Can estimate prevalence from longitudinal or cross-sectional studies
- Generally include all patients in calculation

# Randomised **controlled** trials (RCTs)

---

- **Experimental** and **longitudinal**
- **Comparative** – comparison of two or more treatment strategies (e.g. new regimen vs. existing regimen)
- Control group allows us to conclude that any improvement in outcome is due to the test treatment rather than some other factor
- Where no existing regimen exists, control group may consist of untreated patients (usually receive a **placebo**)

# Randomised controlled trials (RCTs)

---

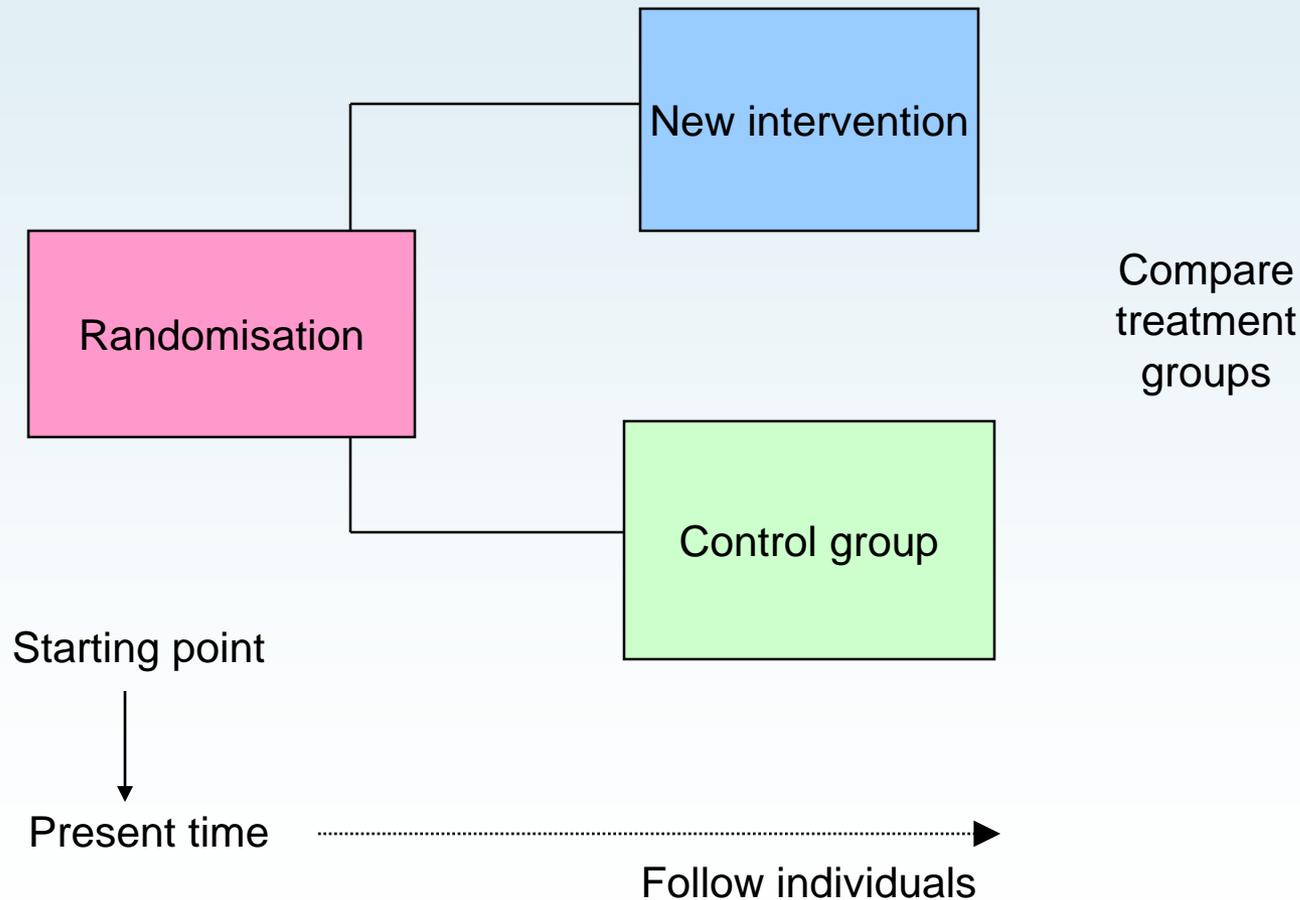
- Subjects allocated to treatment groups by process known as **randomisation**
- Ensures that treatment groups are similar at start of trial; any differences are due to chance only
- Randomisation is most important feature of a RCT and is why RCTs are perceived to be the gold-standard approach to obtaining evidence of a treatment effect
- If you can randomise you should – however, randomisation is not always possible or feasible

# Types of RCTs

---

- **Parallel group:** each patient is randomised to receive only one of the two different strategies
- **Crossover trial:** each patient receives first one treatment strategy then the other, but the treatment order is randomised
- **Cluster randomised:** each 'cluster' of patients (GP surgeries, outpatient clinics) randomised to receive one of the two different treatment strategies

# Parallel design trials

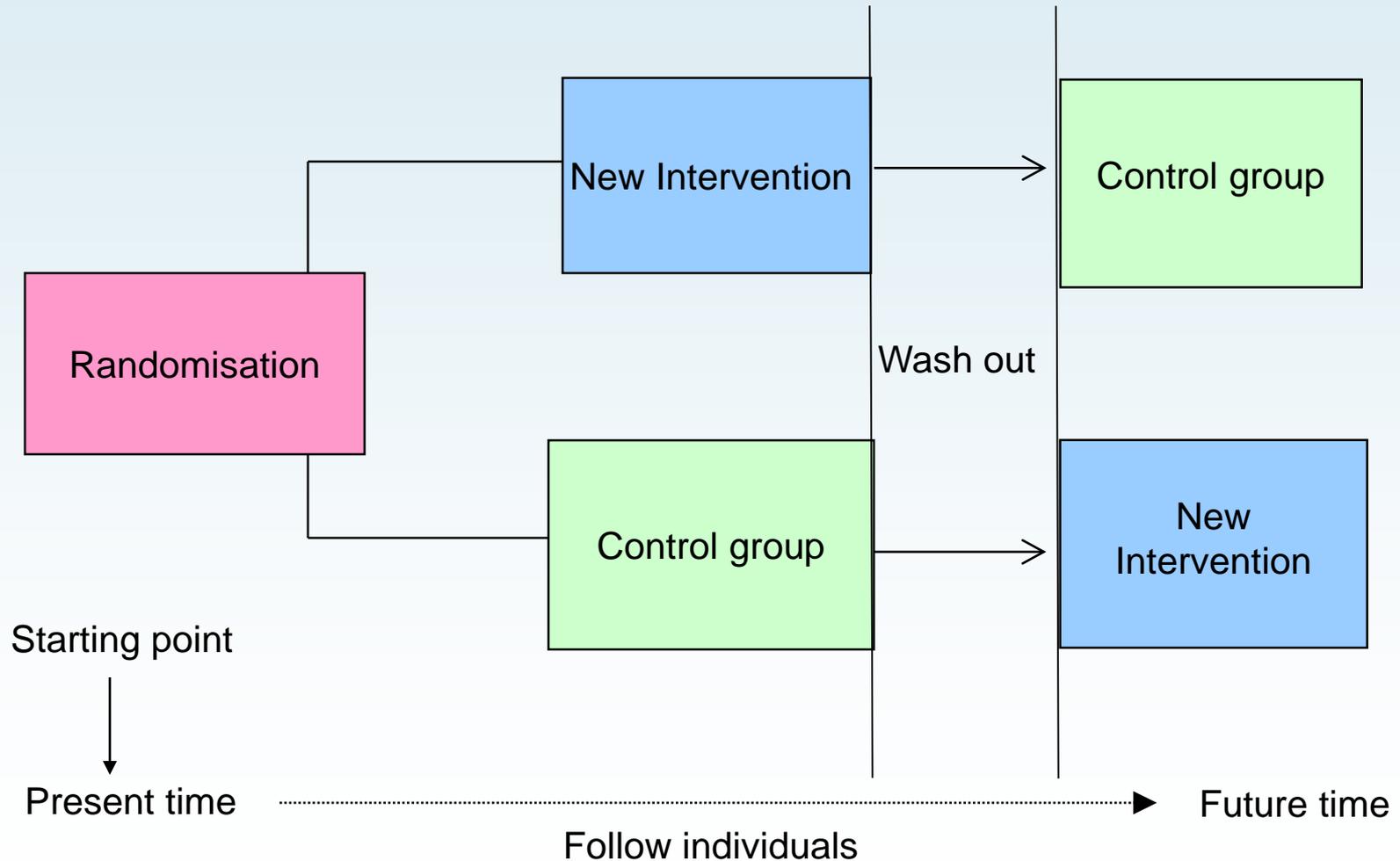


# Example – Parallel Group trial

---

- Trial evaluating when to start ART among HIV-positive individuals who are ART-naïve with CD4 count  $>500$  cells/mm<sup>3</sup>
- Randomised to:
  - Initiate ART immediately following randomisation
  - OR
  - Defer ART until CD4 count is  $<350$  cells/mm<sup>3</sup> or AIDS develops
- Endpoints: Serious AIDS, death from AIDS, serious non-AIDS and death not attributable to AIDS

# Cross-over trials



# Example – Crossover trial

---

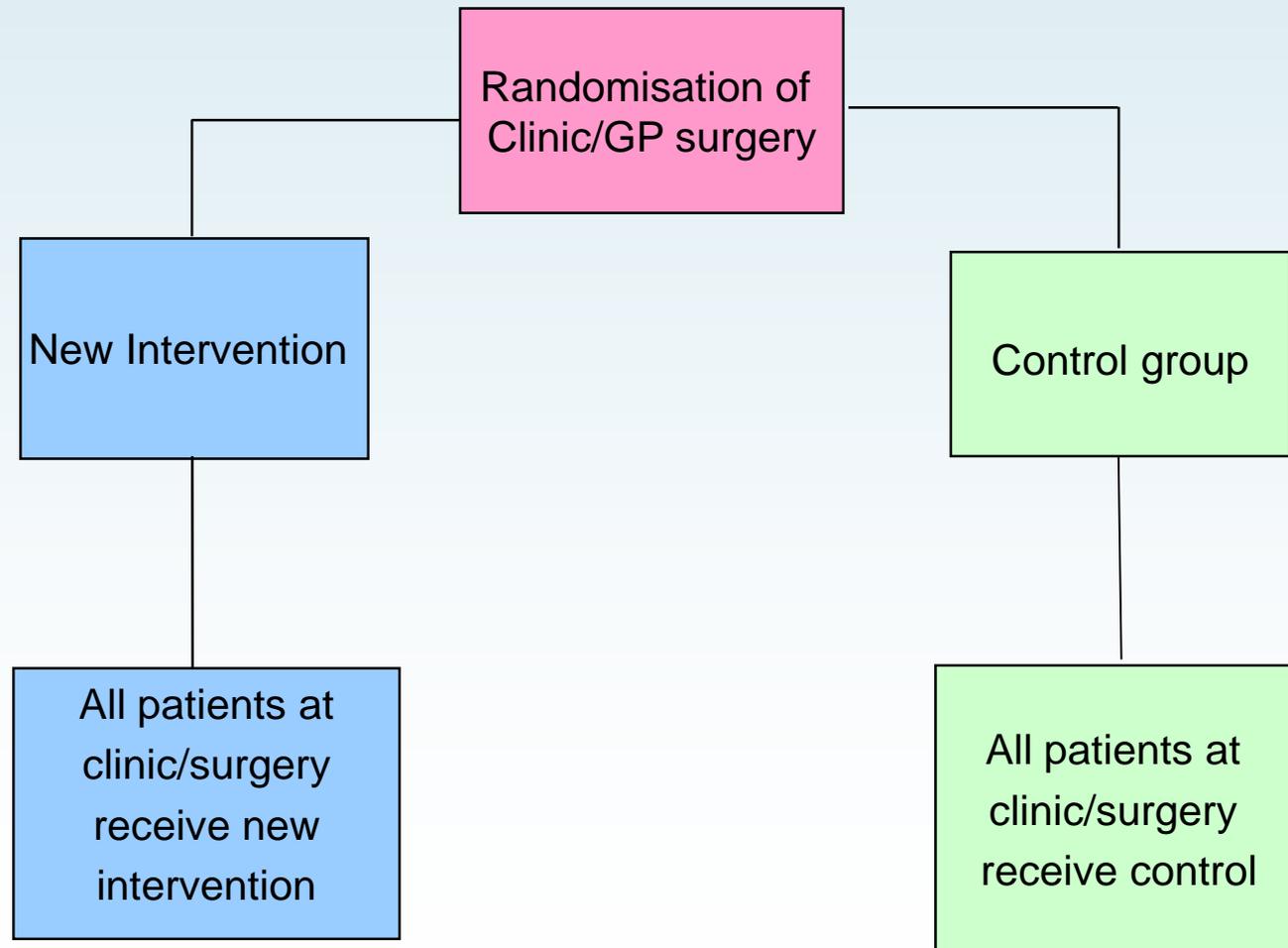
- Safety and acceptability of Reality condom for MSM
- Sero-concordant couples randomised to:
  - Reality condoms for 6 weeks then latex condoms for 6 weeks
  - OR
  - Latex condoms for 6 weeks then Reality condoms for 6 weeks
- Endpoints: frequency of slippage with removal, pain or discomfort on use, rectal bleeding, willingness to use in future

# Crossover trial

---

- Crossover trials are particularly useful for short term outcomes in chronic conditions
- The treatment must be one that does not permanently alter the disease or condition under study
- The main limitation of a crossover trial is that the effect of the first treatment administered may carry over and alter subsequent responses

# Cluster randomised trials



# Example – Cluster randomised trial

---

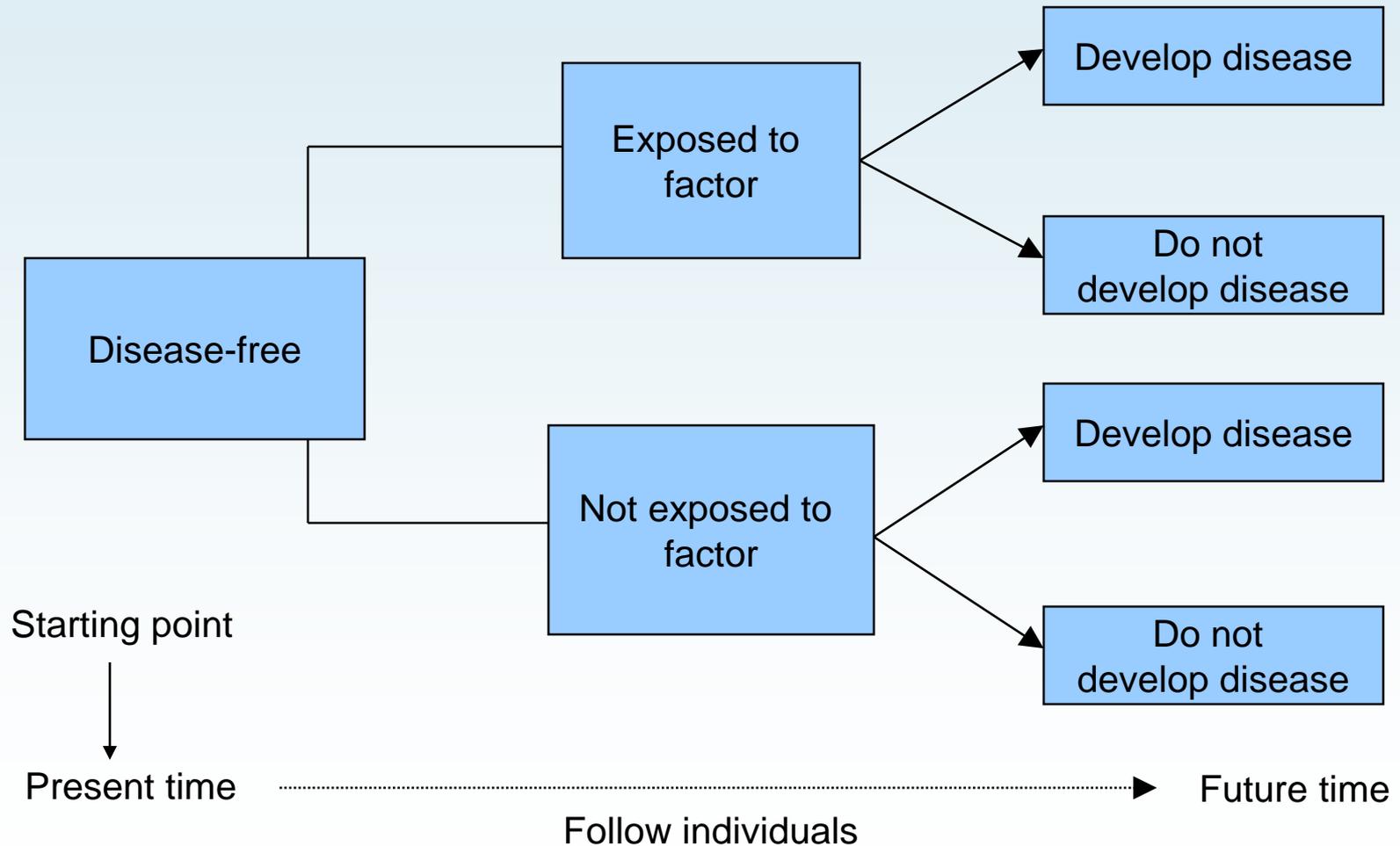
- RCT of malaria prevention in Gambia
- 70 villages randomised to:
  - Long lasting insecticidal nets (LLIN)
  - OR
  - LLIN + indoor residual spraying
- Endpoints:
  - incidence of clinical malaria assessed by passive case detection in >7,000 children
  - number of *Anopheles gambiae* sensu lato mosquitoes collected per light trap per night

# Cohort studies

---

- **Observational** and **longitudinal**
- Follow a group of individuals over time to assess the **incidence** of a disease (or some other outcome)
- Can look at the effect of exposure to a number of factors of interest (potential risk factors) on the incidence of the outcome

# Cohort studies



# Open vs Closed

---

- Closed/Fixed cohorts
  - New patients unable to join study
  - Participant population is fixed at baseline.
  - People can only exit study (withdrawal, death)
- Open/Dynamic cohorts
  - People move in and out of the study.
  - New patients able to join

# Traditional interval cohort

---

- Patients often seen at a study site (often different to their place of care) on regular occasions for ‘study visits’ (e.g. 6-monthly)
- Participants may complete questionnaire on their health since last visit, treatments received, etc.
- Laboratory tests performed at pre-defined time intervals – this information is unlikely to be available at intervening times or when an event occurs, unless this coincides with a study visit
- Patients must give consent to participate

# Traditional interval cohort - example

---

- The Multicenter AIDS Cohort Study (MACS)
- HIV+ve and HIV-ve individuals from 4 centres in Baltimore, Chicago, Los Angeles and Pittsburgh
- Participants recruited from 1984-1985 (n=4954), 1987-1991 (n=668) and 2001-2003 (n=1351)
- Visits are bi-annual – at each visit, participants undergo a detailed interview, physical examination, quality of life assessment and collection of blood for concomitant laboratory testing and storage

# Observational databases

---

- Utilise data collected as part of patient's medical care
- Patient does not attend for a particular study visit
- Laboratory testing performed according to clinical need – will be more frequent if patient is ill or requires investigation
- Some data items may be difficult to collect if not part of routine care
- May or may not require patient consent
- Increasingly common with emergence of electronic record systems

# Observational databases - example

---

- French Hospitals Database on HIV (FHDH)
- Hospital-based multicentre open cohort with inclusions since 1989
- Information on >120,000 patients (53% of French HIV+ population in care)
- Standardized variables collected at each outpatient visit/hospital admission (clinical conditions, treatments prescribed, laboratory tests) and/or at least every 6 months

# Pros and cons of cohort studies

---

## Advantages

- Can assess **temporal relationship** between exposure and disease (i.e. we know which event occurs first)
- Can make some attempt to assess cause and effect

## Disadvantages

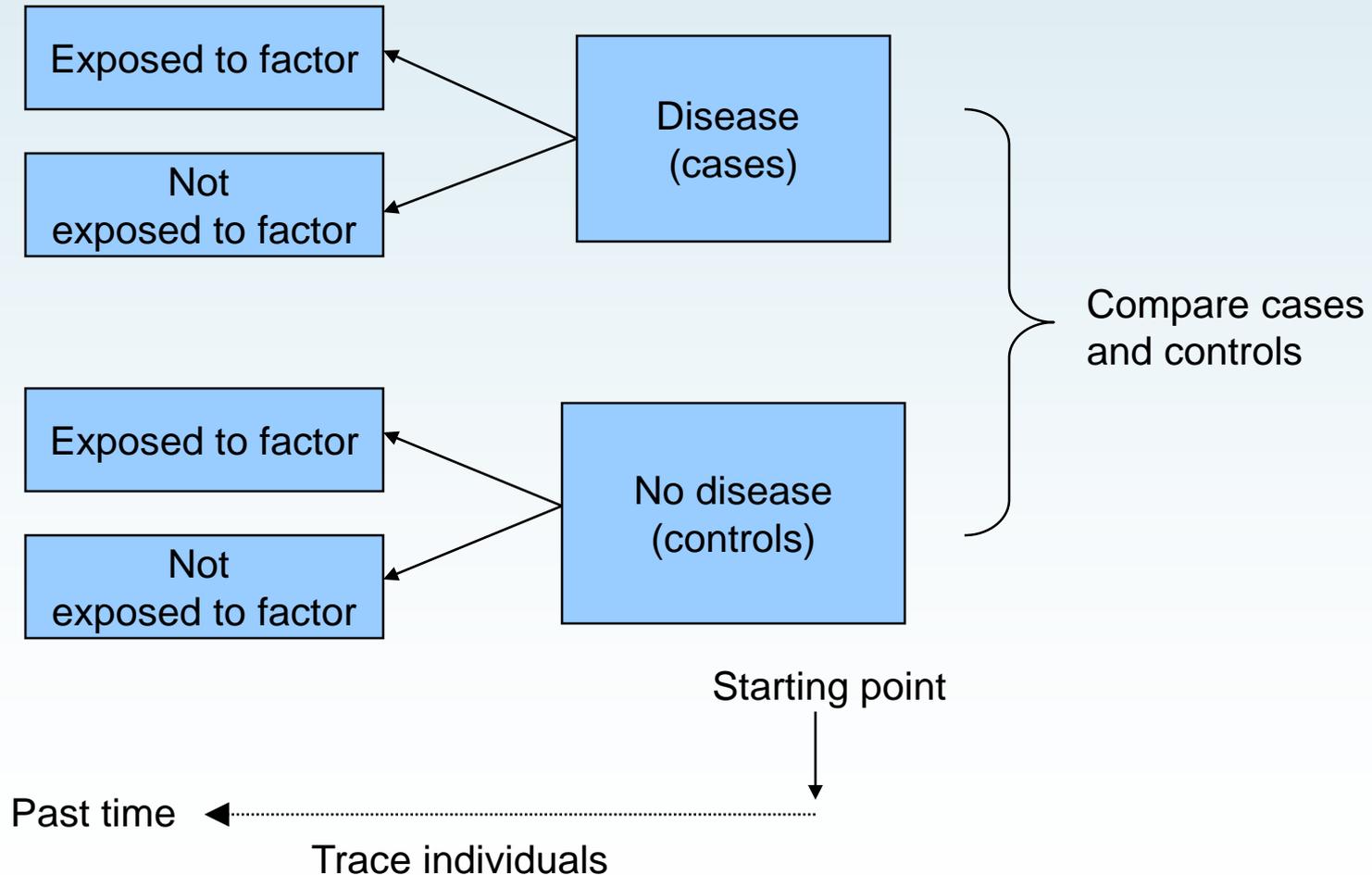
- If the disease is rare then cohort may have to be very large and follow-up long (i.e. expensive)
- May be problem with **loss-to-follow-up**
- Potential for bias due to **confounding**

# Case-control studies

---

- **Observational** and **longitudinal** (retrospective)
- Group of patients with a disease (cases) are compared to group of patients without the disease (controls)
- Aim: has exposure to any factor occurred more or less frequently **in the past** in cases than in controls?
- Cases and controls may often be **matched** on basic demographic information (e.g. sex and age) to make the two groups as similar as possible

# Case-control studies



# Pros and cons of case-control studies

---

## Advantages

- Relatively cheap, quick and easy to conduct
- No loss-to-follow-up
- Suitable for rare events

## Disadvantages

- Potential for **recall bias**
- Timing of events cannot be reliably established – therefore more difficult to assess causality
- Cannot assess incidence (proportion with disease is fixed as part of the study design)

# Predictive factors for HIV seroconversion

---

**Cases:** Persons attending a Spanish HIV unit who seroconverted to HIV >3 months after their first visit following a specific risk of HIV (n=69)

**Controls:** Persons attending same unit after a risk of HIV who did not seroconvert, matched by gender, birthdate and date (n=69)

**Variables:** Demographics, serostatus of partner, exposure risk, previous PEP and STI, PEP regimen, previous HIV testing and presence of STI at baseline

**Conclusions:** Being MSM, having had previous PEP, an HIV-positive sexual partner and previous STI were all predictive factors for HIV seroconversion

# Cross-sectional studies

---

- Carried out at a **single point in time** – no follow-up
- Often used to assess the prevalence of a condition, to describe the current situation or to assess attitudes and beliefs
- Advantages – relatively cheap and quick
- Disadvantages – not possible to estimate incidence of disease, but can assess prevalence

# Alcohol use in HIV+ve persons

---

- Cross-sectional study: 2230 HIV+ve patients in 3 primary care clinics in Pretoria
  - - 25.1% reported hazardous or harmful drinking
  - - 2.0% had possible alcohol dependence
- In multivariable analyses, high-risk drinking associated with male gender, never being married, tobacco use, greater independence and more depressive symptoms
- Recommendation of routine screening for alcohol use and harm reduction interventions

# Case series / case-note review

---

- Fairly low form of evidence but can provide useful preliminary data
- Useful as a descriptive tool – i.e. to define the natural history of disease or to describe current practices
- No comparative element – therefore not possible to show a link between exposure and disease
- Usually retrospective – therefore potential for problems with historical data

# Choosing an appropriate study design

---

- The hypotheses that can be tested in any study, particularly regarding ‘cause and effect’, will depend on the study design
- Some study designs may offer ‘benefits’ in terms of cost, time and administrative effort, but in general, studies that are quicker and cheaper to perform will provide weaker evidence
- Must have a clear idea of the hypotheses being tested before choosing the optimal study design

# Research question

---

## **QUESTION:**

**Do elderly (>70 years), Irish, female patients with metabolic syndrome and first presentation of TIA who have standard, multi-specialist (endocrinology, cardiovascular, gerontology) care have higher one-year mortality compared to those receiving integrated (endocrinology, cardiovascular, gerontology) guideline-driven, single centre specialist care within a metabolic clinic?**

# Summary

---

- The hypotheses that can be tested in any study, particularly regarding ‘cause and effect’, will depend on the study design
- Some study designs may ‘offer’ benefits in terms of cost, time and administrative effort – these are likely to provide weaker evidence
- All studies involve the selection of a sample – if the sample is not representative, the results of the study may be biased