EACS HIV Summer School 2016

Mini lecture 3:

Conducting and managing observational studies

Colette Smith

UCL Research Department of Infection and Population Health

5th September 2016

Outline of Session

- The limitations of RCTs
- Designing a cohort study
- Designing a case-control study

Outline of Session

- The limitations of RCTs
- Designing a cohort study
- Designing a case-control study

The key features of RCTs

- The control group allows us to judge whether a new treatment provides additional benefits to those expected with standard care
- Randomisation and blinding (where used) allow us to conclude that any such benefits are due to the treatment itself, rather than to any other factor or to chance
- As a result, RCTs are perceived to be the gold-standard method for obtaining evidence of a treatment effect

However, RCTs may have some limitations

1. RCTs are only possible where there is an ‘intervention’ that people are willing to be randomised to

However, RCTs may have some limitations

1. RCTs are only possible where there is an 'intervention' that people are willing to be randomised to

   Examples…

   - Impact of smoking and/or alcohol consumption on response to HAART
   - Impact of co-infection with TB on HIV progression rates

However, RCTs may have some limitations

2. Patients in RCTs may not be representative of the clinic population, and follow-up of patients may differ to that in clinic – thus, outcomes may differ from what would normally be expected

Examples…

- Patients may be selected on the basis of their likely adherence to treatment
- Patients may attend clinic more frequently – outcomes may be detected sooner
- Monitoring may be more intensive

However, RCTs may have some limitations

3. RCTs may be short (48/96 weeks) and may focus on two or three main treatment comparisons

However, RCTs may have some limitations

4. RCTs may concentrate on short-term surrogate marker endpoints rather than long-term clinical events

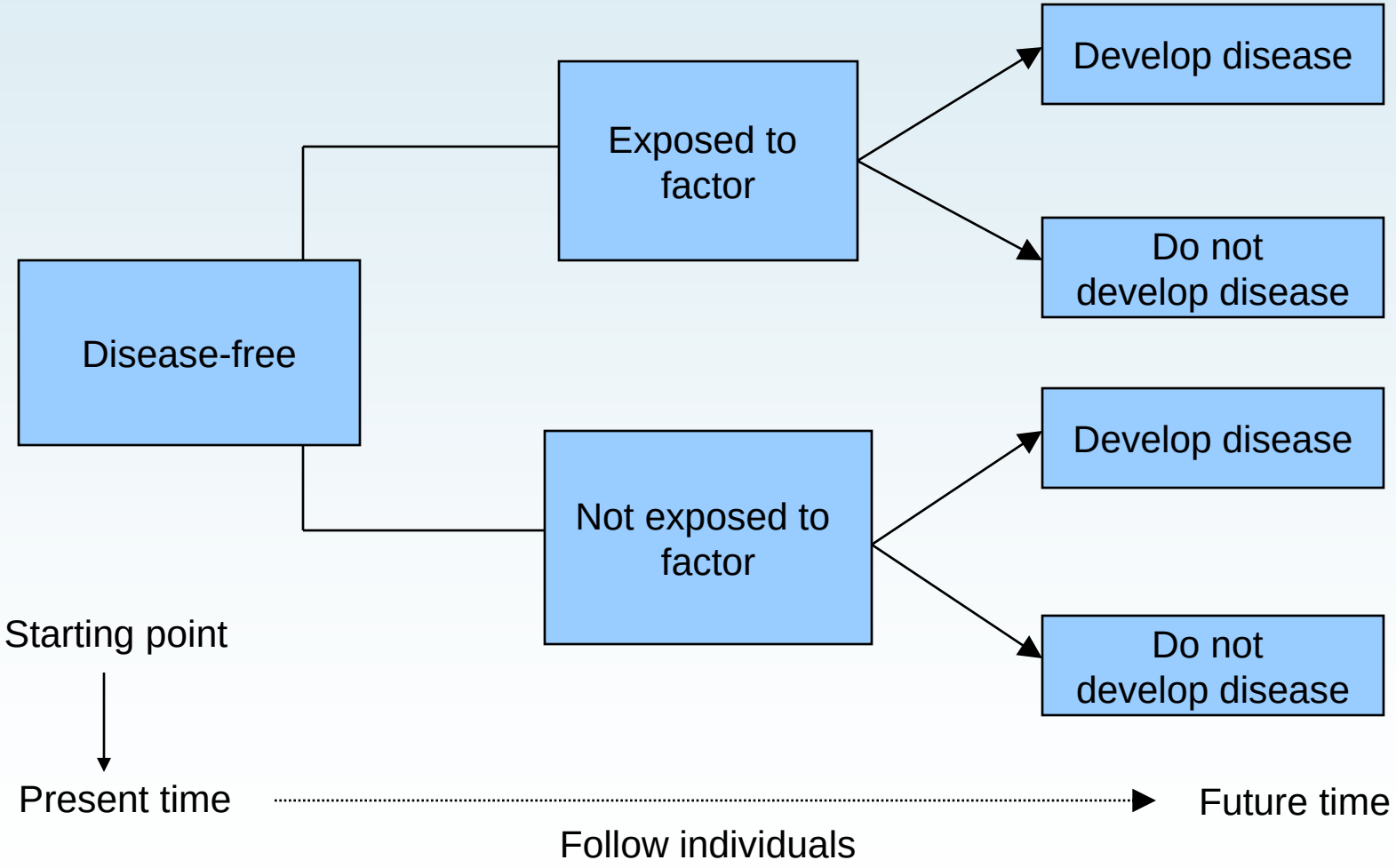
   Example…

   - Early studies of IL-2 treatment in HIV infection focussed on CD4 endpoints only

Outline of Session

- The limitations of RCTs
- Designing a cohort study
- Designing a case-control study

Reminder – Cohort Studies



Basic study design issues

- Important to have a clear objective for the study and to design accordingly – set out in a protocol
- Ensure that sample size will be sufficient to address at least one key hypothesis
- Participants included in cohort should be representative of the population to which the results will be generalised
- Try to collect data on all variables known to predict the disease outcome of interest (especially confounders)

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
- Patients may be seen at regular time intervals for specific 'cohort' visits or the cohort may make use of existing data collection systems within clinics

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 under go a detailed interview, physical examination, quality of life assessment and collection of blood for concomitant laboratory testing and storage

Observational databases

- Utilise data routinely 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
- Currently has information >120,000 patients (53% of French HIV+ population in care; representative in terms of geographical origin, sex, age)
- Standardized variables collected at each outpatient visit/hospital admission at which new clinical condition diagnosed, new treatment prescribed, laboratory test performed and/or at least every 6 months

Mary-Krause et al; Int J Epidemiol; 2014; 43(5): 1425-1436

Distinction between the two types of cohort

	Traditional cohort	Observational database
Study visits	At regular defined intervals	As and when patient attends for care
Data entry	Often form-based	Often electronic transfer of data
Representative?	May not be – patients must give consent	Sometimes includes *all* patients
Loss to follow-up	May be substantial, but can be determined	May be difficult to assess as some patients attend infrequently
Data quality	Can introduce quality control measures	Difficult to regulate
Data items collected	Can determine at outset and change over time	May be difficult to influence

Representativeness

- Validity of study results relies on sample being representative
- This is influenced by:
  - Selection of study sample
  - Response rate
  - Poor measurement of exposure & outcome
  - Loss to follow-up - a significant challenge for longitudinal studies

Loss to follow up

- People who drop out may be more likely to:
  - Have poorer physical functioning/greater risk of experiencing outcome
  - Live alone
  - Lower socio-economic status
- Why may people leave studies?
  - Move house
  - Too time-consuming
  - Drop out of medical care
  - Travel to clinic difficult
  - Dislike of medical tests
  - Concerns for data confidentiality

Ways of following individuals

- Failure to ascertain all disease events, including due to loss to follow-up, can result in under-estimation of event rates
- Can also lead to bias in comparisons between exposure levels
- Nationally recorded registers/databases
  - Death certificates
  - Disease registers, e.g. cancer registry
  - (In UK) NHS electronic information systems e.g. hospital episode statistics, GP databases
  - Office of national statistics (ONS)
- Other efforts to contact people (e.g. phone call, letter)

Key outcome variables

- An ideal outcome should address the primary aim of the study, have biological/clinical relevance and be appropriate for the population studied
- It should be ascertainable on all cohort participants (including those lost-to-follow-up)
- In the HIV setting, common outcomes include:
  - **Clinical:** New AIDS events, new non-AIDS event, death
  - **Virological:** VL<50 copies/ml at 1 year, time to viral suppression, time to viral rebound,
  - **Immunological:** CD4>200 cells/mm^3, time to CD4 increase >100 cells/mm^3
  - **Other:** ART switches, adherence, quality of life, toxicity

Toxicity outcomes

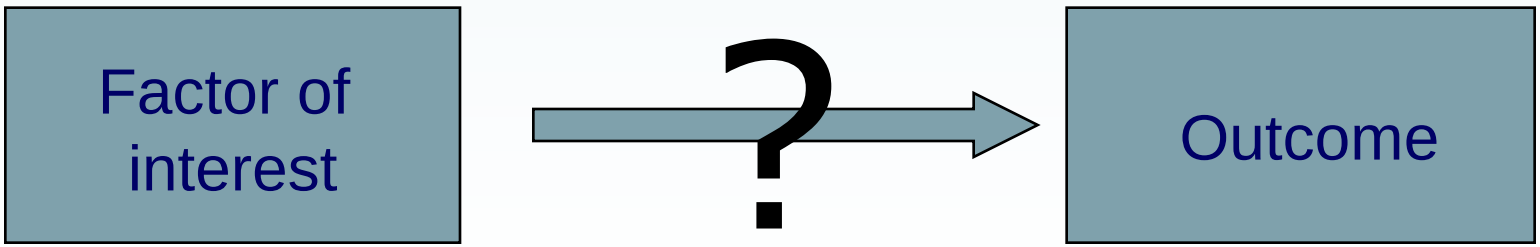
- Cohorts may be the only study design that is able to capture data on long-term toxicities of HAART
- Toxicity outcomes may be based on clinical symptoms and/or laboratory data
- Need to be aware of possible biases when interpreting results from such studies:
  - Irregular/infrequent laboratory monitoring
  - Selective laboratory monitoring
  - Between-clinic assay variability
  - Clinic differences in monitoring policies
  - Bias due to confounding

What other data should be captured?

- Potential confounders (traditional definition):
  - Factors that are associated both with the exposure and outcome of interest
  - Failure to adjust for confounders may introduce bias, as they may lead to a spurious association between the exposure and outcome
- Effect modifiers:
  - Factors that modify the size of the association in one group compared to another
  - Provide important clinical information
  - Often referred to as a ‘statistical interaction’

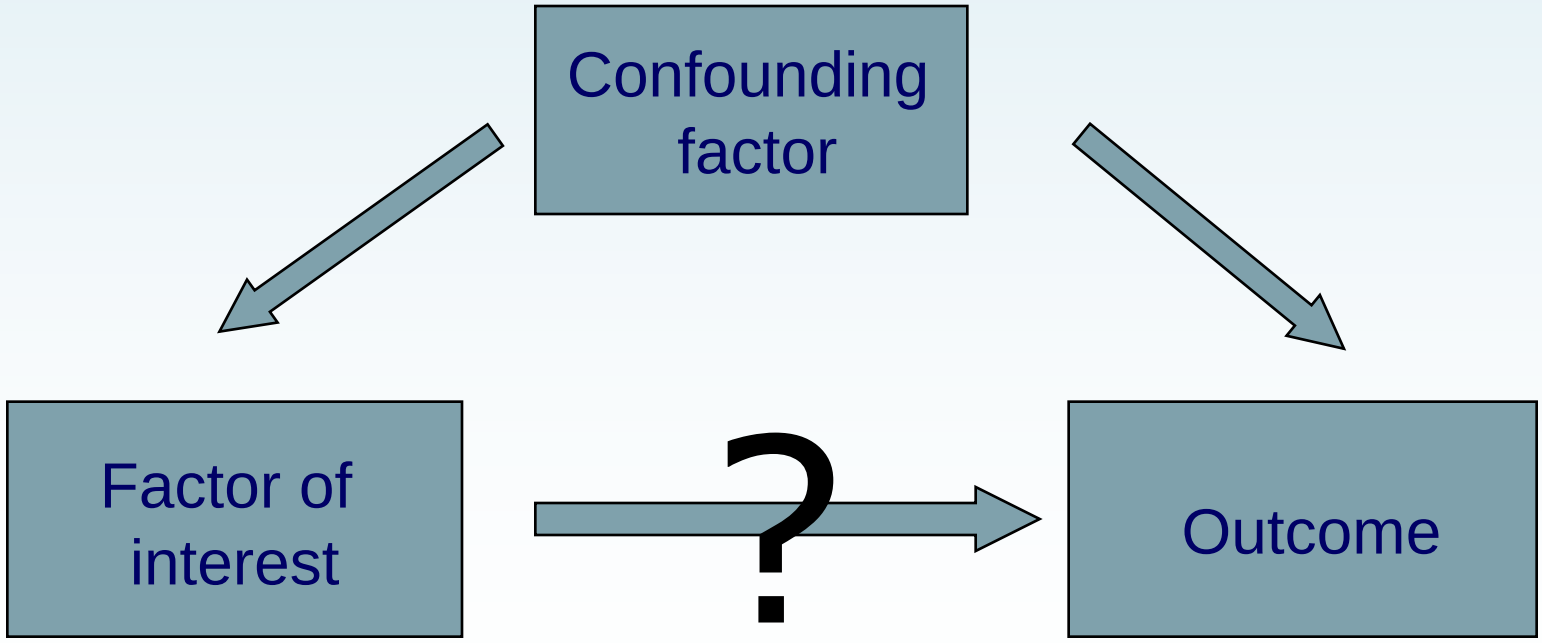
Bias due to confounding

- Occurs when a spurious association arises due to a failure to fully adjust for factors related to both the risk factor and outcome.



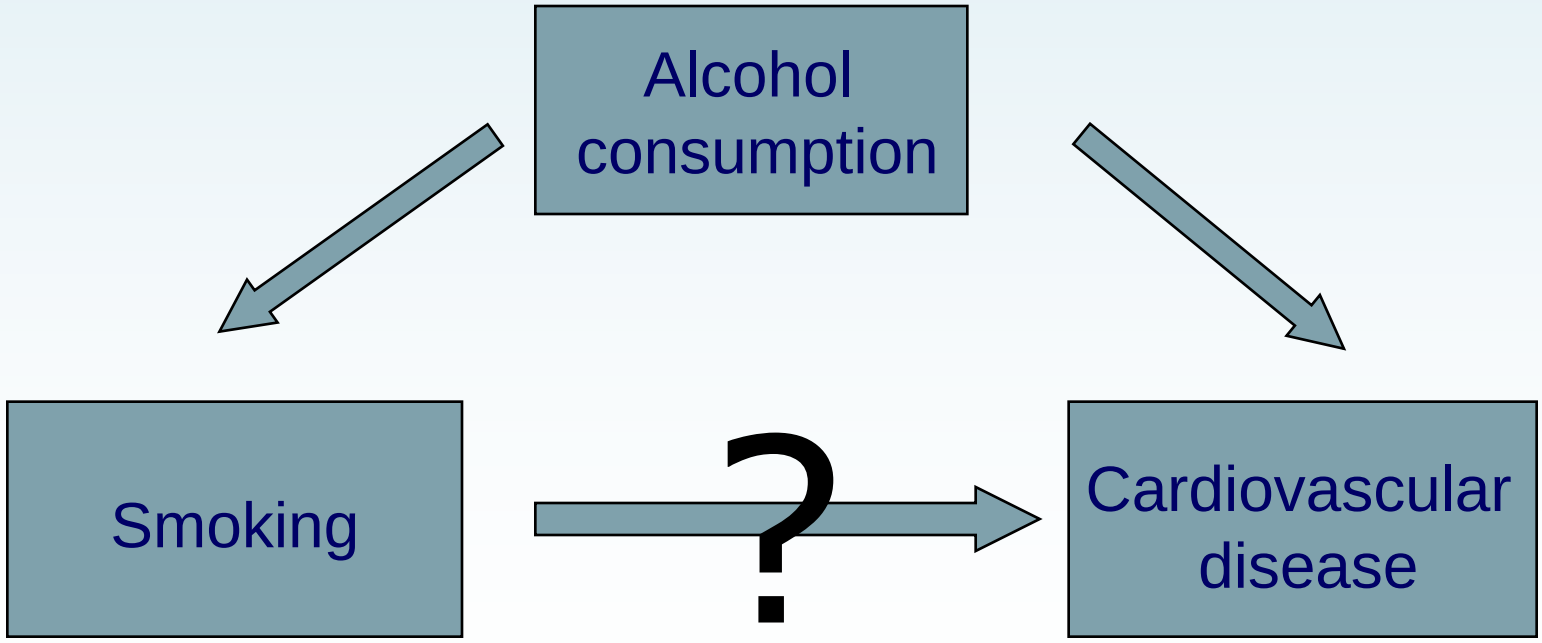
Bias due to confounding

- Occurs when a spurious association arises due to a failure to fully adjust for factors related to both the risk factor and outcome.



Bias due to confounding

- Occurs when a spurious association arises due to a failure to fully adjust for factors related to both the risk factor and outcome.



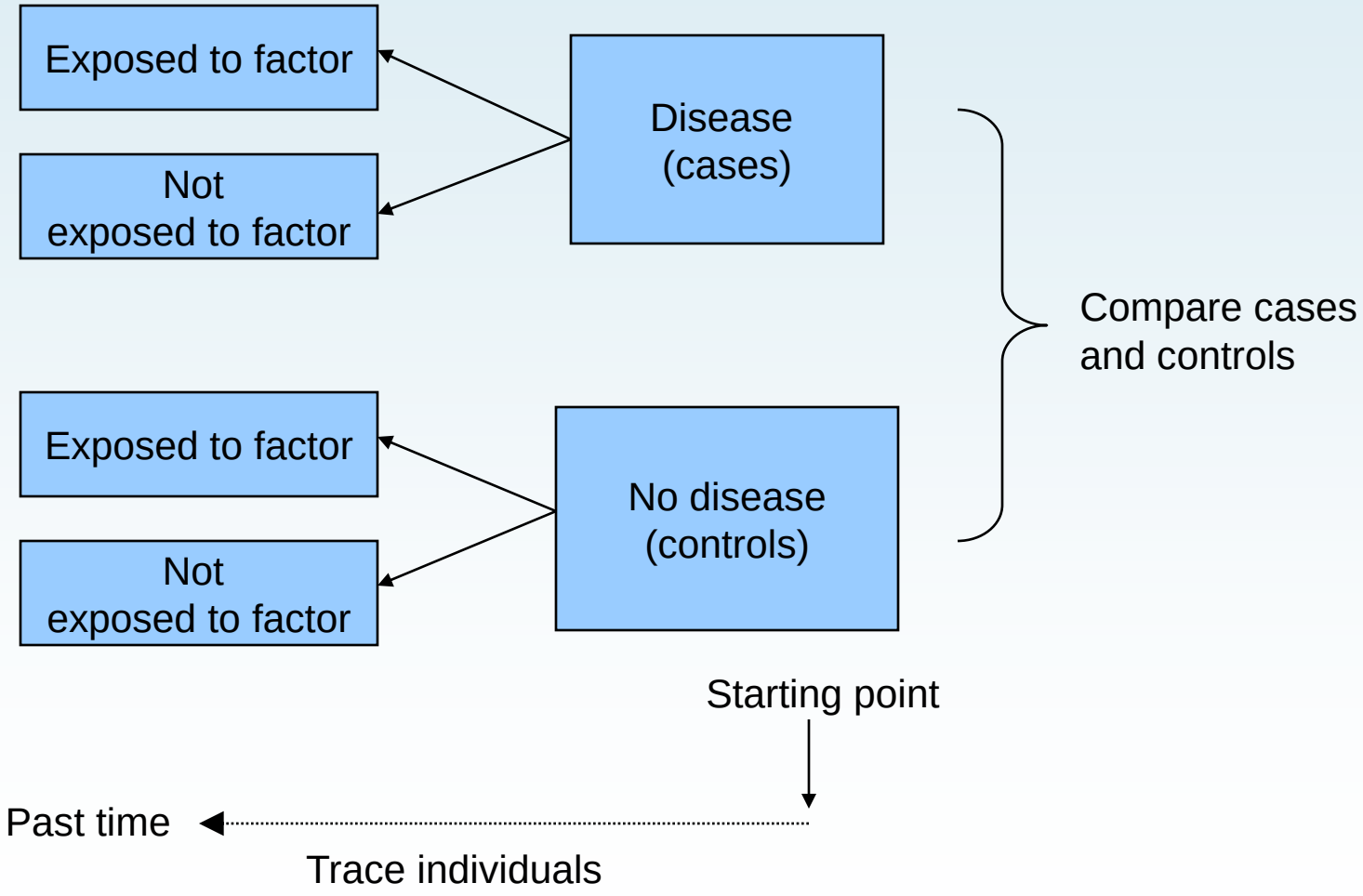
Exposures, confounders and effect modifiers

- As measurements may change over the study period, a patient's status should be re-assessed at regular times during the study
- The frequency at which each measurement is assessed will depend on the likelihood of it changing over time, as well as the reliability of the data sources
- Example: dietary factors, smoking status, alcohol consumption

Outline of Session

- The limitations of RCTs
- Observational studies – their benefits and limitations
- Designing a cohort study
- Designing a case-control study

Reminder – Case-control studies



General points

- Retrospective, so reliant on recorded data (which may contain inaccuracies and be subject to missing data)
- Useful for rare diseases and diseases with long latency periods
- Care should always be taken to ensure that the timing of events (e.g. exposures, outcome) is captured accurately

Selection of cases

- Develop a **case definition** to identify cases that is precise, objective and unambiguous
- This could include:
  - Histologically or laboratory confirmed diagnosis
  - Clinical diagnosis
  - Stages of disease (standardised e.g. CDC AIDS definition)
- Source of cases needs to be carefully considered
  - Population based or clinic based cases?
  - How complete is your source of cases?
  - Is there a time lag between diagnosis and notification/ identification
  - What about patients who may have moved or died

Incident or prevalent cases

- **Incident cases (best)**: All **new** cases identified within the study population over a specified time.
  - Better recall of exposure history
  - Less likely to have altered behaviours (risk factors) in response to diagnosis.
  - More expensive in terms of time and resources

- **Prevalent cases**: All cases (**new and existing**) within the study population who had the disease at a particular time point
  - Less expensive
  - Useful for conditions where date of onset difficult to establish
  - Patients with long duration of disease over represented
  - Survival bias

Selection of controls

- Controls should be selected to be as similar to cases as possible except for the outcome of interest
  - Drawn from the same population
  - Fulfil the same eligibility criteria
- Purpose is to provide an estimate of "level of exposure" in those without outcome
- Should represent population from which the cases are drawn
  - General population (voting registries, random digit dialling etc)
  - Hospital/clinic based controls – care should be taken to ensure they don't have another condition also related to the exposure
  - People related to the case – i.e. friends, relatives, neighbours

Matching in case-control studies

- Cases and controls may often be matched on a small number of factors associated with both the exposure and outcome (e.g. sex, age)
- Matching may help to minimise effects of confounding and may increase study power
- However, it may be impractical to match patients on many factors and special analytical methods may be required if matching is used
- If a factor has been used in matching, then it is not possible to evaluate its association with the outcome

Recall bias

- Tendency of cases to ‘recall’ information (particularly relating to exposure) differently to cases
- Can lead to apparent association between outcome and exposure, even if the association does not exist
- Example: cigarette smoking and lung cancer

Example – HIV in health care workers

- Association between receipt of ZDV post-exposure prophylaxis and risk of HIV acquisition
- **Eligibility criteria:** HCWs in USA with documented exposure to HIV blood through needle stick injury between 1988 – 1994
- **Case definition:** HCWs who experienced HIV seroconversion associated with the exposure, and no other concurrent exposure to HIV (n=31)
- **Control definition:** HCWs who remained HIV negative at exposure & at least 6m later (n=679)

http://www.cdc.gov/mmwr/preview/mmwrhtml/00039830.htm#00001426.htm

Example – HIV in health care workers

- **Case selection:** identified through reports to national surveillance systems for occupationally acquired HIV infection operated by CDC
- **Control selection:** Identified through reports to a passive surveillance project maintained by CDC since 1983 (from ~300 health-care institutions in USA)
- **Exposure of interest:** PEP to ZDV identified through reports to a passive surveillance project maintained by CDC since 1983 (from ~300 health-care institutions in USA)

Nested case-control studies

- A case-control study may often be nested within a larger cohort or RCT
- This provides a means of studying associations between novel biomarkers and disease outcome, particularly if these are expensive to measure
- Alternatively, may be useful if additional detailed information is required which cannot be collected through standard data collection mechanism
- Example: nested case-control study in SMART trial, measured lipoprotein particles in 248 patients with a CVD event (cases) and 480 matched controls

Which study design?

- Largely an issue of resources, although it is possible to use a combination of approaches
- Important to consider whether the data collected using a particular design will be able to answer the question of interest
- Can often be a compromise between costs and the amount and type of data that can be collected
- Ultimately it is important to be aware of, and transparent about, the limitations of each study design

Where to go for guidance?

- Similar to CONSORT but provides guidance on reporting of observational studies
- Provides a checklist for reporting studies, as well as educational material
- Recommendations limited to 3 main designs of observational studies
  - Cohort
  - Case-control
  - Cross-sectional studies
- www.strobe-statement.org

Summary

- Whilst RCTs are perceived to provide the highest quality evidence when assessing associations, they may sometimes suffer from limitations which make them inappropriate for use when addressing certain questions
- In these situations, observational studies may provide useful information
- However, observational studies are always subject to bias and must be designed, managed and interpreted with caution so as to minimise this

Topics for this year

- Co-infections / Co-morbidities
- Long term management
- Cascade of care
- Diagnosis and initiation of treatment

Organisation of working groups : Monday

- Select one topic within the main orientation of the group for further development (60 minutes*)
- Develop topic into a formal research question (30 minutes*)
- Discuss pros and cons of different study designs (30 minutes*)

* For guidance only

Organisation of working groups : Tuesday

- Prepare protocol (120 minutes*)
  - Identify the appropriate study population
  - Identify and define the key outcomes
  - Define the intervention (or key exposure variables)
  - Identify potential confounders (if applicable)
  - Determine approach for enrolling and following study subjects

* For guidance only

Organisation of working groups : Thursday

- Sample size and other statistical issues
- Prepare presentation
  - 1 presenter for each subgroup

* For guidance only