MONASH University
Medicine, Nursing and Health Sciences

Case Control studies

Malcolm Sim

Centre for Occupational and Environmental Health
School of Public Health and Preventive Medicine
Monash University

Summary of Talk

- Epidemiological study designs
- Features of case control studies
- Selecting cases and controls
- Measuring disease and outcome
- Calculating an odds ratio
- Nested studies
- Power considerations
- Advantages and disadvantages

Epidemiological Study Designs

- Intervention Studies - explore the association between interventions and outcomes. (Experimental studies or clinical trials) Unusual in occ cancer research
- Observational Studies - examine associations between risk factors and outcomes (Analytical - determinants and risk of disease, and descriptive - patterns and frequency of disease)

Epidemiological study designs

True epidemiological designs

- Randomised controlled trial
- Intervention trial
- Cohort study
- Case cohort study
- Case control study
- Cross-sectional study
- Ecological study

Others

- Case report
- Case series
- Retrospective chart reviews
- (No controls)

Case-Control Studies

- Another type of analytical observational study
- Unit of observation and analysis: Individual (not group), as able to obtain individual level information
- ‘Simple’ design and cost effective to conduct

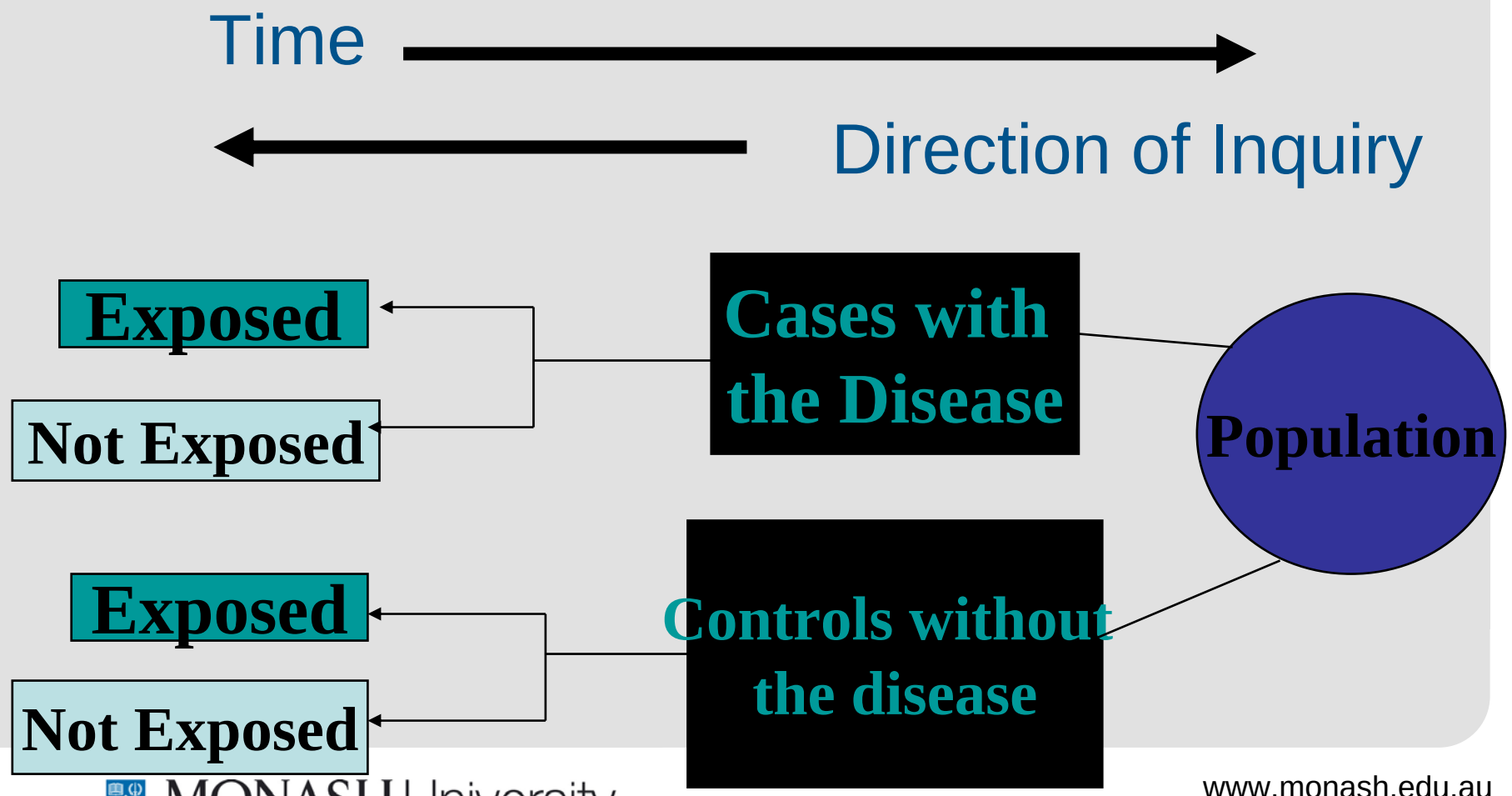
Case-Control Studies

- Case-control studies are the most frequently undertaken analytical epidemiological studies
- They are the only practical approach for identifying risk factors for rare diseases
- Can investigate many different exposures
- They are best suited to the study of diseases for which medical care is sought, such as cancers or other chronic conditions

Overall study design

- At baseline:
  - **Selection of cases (disease) and controls (no disease) based on disease status**
  - **Exposure status is unknown**
- Retrospective design – assess exposure in the past

Case Control Design (retrospective enquiry)



Selecting Cases

- Select cases after the diagnostic criteria and definition of the disease are clearly established
- Study cases should be representative of all cases, so no response bias

Selecting Cases (cont 2.)

- The study need not include all cases in the population – power considerations
- Cases may be identified from hospitals, clinics, disease registries, screenings, etc.

Selecting Cases (cont 3.)

- Incident cases are preferable to prevalent cases for reducing:
  - **(a) recall bias and**
  - **(b) over-representation of cases of long duration, as may be different from those who die early**
- The most desirable way to obtain cases is to include all incident cases in a defined population over a specified period of time

Selecting Controls

- One of the most important aspects of case control studies
- Should be as similar as possible to the cases, but without the disease being studied
- Controls should be selected from the same population at risk for the disease as the cases
- Controls should be representative of the target population

Selecting Controls (cont.)

- If using hospital-based controls, need to be careful that the reason that the controls were in hospital is not related to the same risk factors being investigated
- Controls estimate the exposure level to be expected in cases if there were no association between exposure and disease

Selecting Controls (cont.)

- Multiple controls can be used to help add statistical power when disease is uncommon or cases are unduly difficult to obtain
- More than 3 controls for a case is usually not cost-efficient and diminishing increase in power
- Using more than one control group can lend credibility to the results if consistent result, but can be a problem if conflicting results. Also cost implications. Therefore think carefully about which control group is most appropriate up front

Sources of cases and controls

CASES	CONTROLS
All cases diagnosed in the community	Sample of general community
All cases diagnosed in a sample of the population	Non-cases in a sample of the population
All cases diagnosed in all hospitals	Sample of patients in all hospitals who do not have the disease
All cases diagnosed in a single hospital	Sample of patients in the same hospital who do not have the disease
Any of the above methods	Spouses, siblings, neighbours or associates of cases

Assessing Exposure

- Exposure is usually an estimate unless past measurements are available
- It has to be assumed that the exposure occurred prior to the time the disease process began (this may not be valid)
- Usually collected by questionnaire and may be supplemented by records

Assessing Exposure (cont.)

- Questionnaire data should be collected using standardised modules, eg smoking history
- Job calendar useful, then use relevant job specific modules (JSMs) for exposure information.
- Industry/Job/task data can be reviewed using expert hygienist assessment method, singly or as a panel

Assessing Exposure (cont.)

- Exposure estimates are subject to *recall bias* and *interviewer bias*
- Recall bias a major problem. People with the disease more likely to remember past exposures
- Interview bias can occur where interviewers more strongly interrogate cases than controls
- Some protection may be afforded by
  - blinding interviewers (not always possible)
  - carefully phrasing interview questions or including questions about unlikely exposure variables

Potential confounders

- Potential confounders need to be considered in the study design
- Can reduce impact by:
  - Restriction, eg nonsmokers
  - Matching, eg age range
  - Adjustment in the analysis

Matching

Controls can be individually or frequency matched

INDIVIDUAL MATCHING: search for one (or more) controls who have the required MATCHING CRITERIA, such as age, gender.

FREQUENCY MATCHING: select a population of controls such that the characteristics of the group match the characteristics of the cases. e.g. if 15% of cases are males, 15% of the controls are also.

Avoid 'overmatching', where variable related to exposure

Odds Ratio (OR)

- A ratio that measures the odds of exposure for cases compared to controls
- Odds of exposure = number exposed ÷ number unexposed
- OR Numerator: Odds of exposure for cases
- OR Denominator: Odds of exposure for controls

Calculating the Odds Ratio

		Disease Status: Cancer cases (Cases)	Disease Status: No cancer (Controls)
Exposure Status	Exposed	112	176
	Non-exposed	88	224
	Total	200	400

Odds Ratio $= \frac{A/C}{B/D} = \frac{AD}{BC} = \frac{112 \times 224}{176 \times 88} = 1.62$

MORE POINTS ABOUT CASE-CONTROL ANALYSIS

- The odds ratio is a good estimate of the relative risk when the disease is rare (prevalence < 20%).
- Can be extended to N > 1 controls.
- statistical testing is by simple chi-square (unmatched analysis) or by McNemar's chi square (matched-pairs analysis).
- Can be extended to multiple strata (Mantel-Haenzel chi-square)

	OR<1	OR=1	OR>1
Odds comparison between cases and controls	Odds of exposure for cases are less than the odds of exposure for controls	Odds of exposure are equal among cases and controls	Odds of exposure for cases are greater than the odds of exposure for controls
Exposure as a risk factor for the disease?	Exposure reduces disease risk (Protective factor)	Particular exposure is not a risk factor	Exposure increases disease risk (Risk factor)

Interpreting the Odds Ratio

The odds of exposure for cases are 1.62 times the odds of exposure for controls.

or

Interpreting the Odds Ratio

Those with cancer are 1.62 times more likely to be exposed than those without the cancer

or

Those with cancer are 62% more likely to be exposed than those without the cancer

Possible Sources of Bias and Error

- Information on the potential risk factor (exposure) may not be available either from records or the study subjects' memories
- Information on potentially important confounding variables may not be available either from records or the study subjects' memories
- Source of random error when equal for cases and controls, but bias when differential

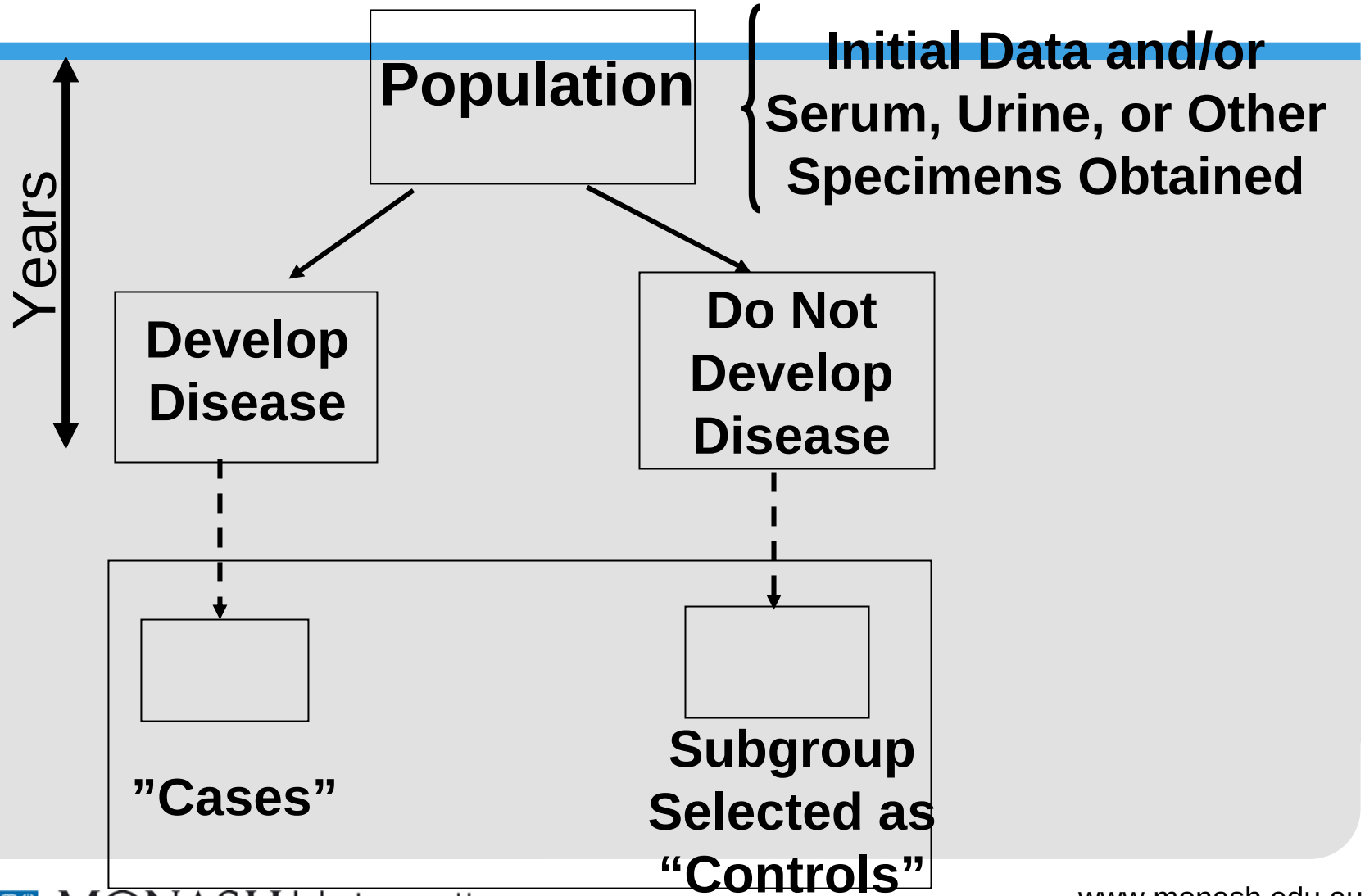
Possible Sources of Bias and Error (cont.)

- Cases may search for a cause for their disease and thereby be more likely to report an exposure than controls (recall bias)
- The investigator may be unable to determine with certainty whether the suspected agent caused the disease or whether the occurrence of the disease caused the person to be exposed to the agent – temporal relationship may be unclear

Possible Sources of Bias and Error (cont.)

- Identifying and assembling a case group representative of all cases may be unduly difficult
- Identifying and assembling an appropriate group of controls may be unduly difficult

Nested Case-Control Study



CASE-CONTROL STUDY

ORs, P-Values and 95% CIs for Case-Control Study with 3 Different Sample Sizes

	Sample Size		
Parameter Computed	n=20	n=50	n=500
OR	**2.0**	**2.0**	**2.0**
p-value	**0.500**	**0.200**	**0.001**
95% CIs	**0.5, 7.7**	**0.9, 4.7**	**1.5, 2.6**

Advantages of Case-Control Studies

- **Quick and easy to complete, cost effective**
- **Cases usually readily available**
- **Most efficient design for rare diseases**
- **Usually requires a smaller study population than a cohort study**
- **Can study many exposures**

Disadvantages of Case-Control Studies

- **Uncertainty of exposure-disease time relationship**
- **Inability to provide a direct estimate of risk. Can't calculate incidence rate.**
- **Not efficient for studying rare exposures**
- **Subject to biases (recall, selection and interviewer bias)**
- **Can only study one outcome**

Disadvantages of Case-Control Studies 2

- **Heterogeneity in case definition if cases from different hospitals**
- **Cases may not be representative of all cases, eg tertiary referral**
- **Cases may be too ill to take part or have died, unless accessed early**
- **Proxy respondents not recommended, especially if differential between cases and controls**