MONASH University
Medicine, Nursing and Health Sciences

Cohort 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 cohort studies
- Assembling a cohort
- Measuring disease and outcome
- Incidence rates and person years
- Nested studies
- Advantages and disadvantages

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)

What is a cohort?

- **Ancient Roman military unit, A band of warriors.**
- **Persons banded together.**
- **Group of persons with a common statistical characteristic. [Latin]**
- **e.g. age, work group, live in a specific area**

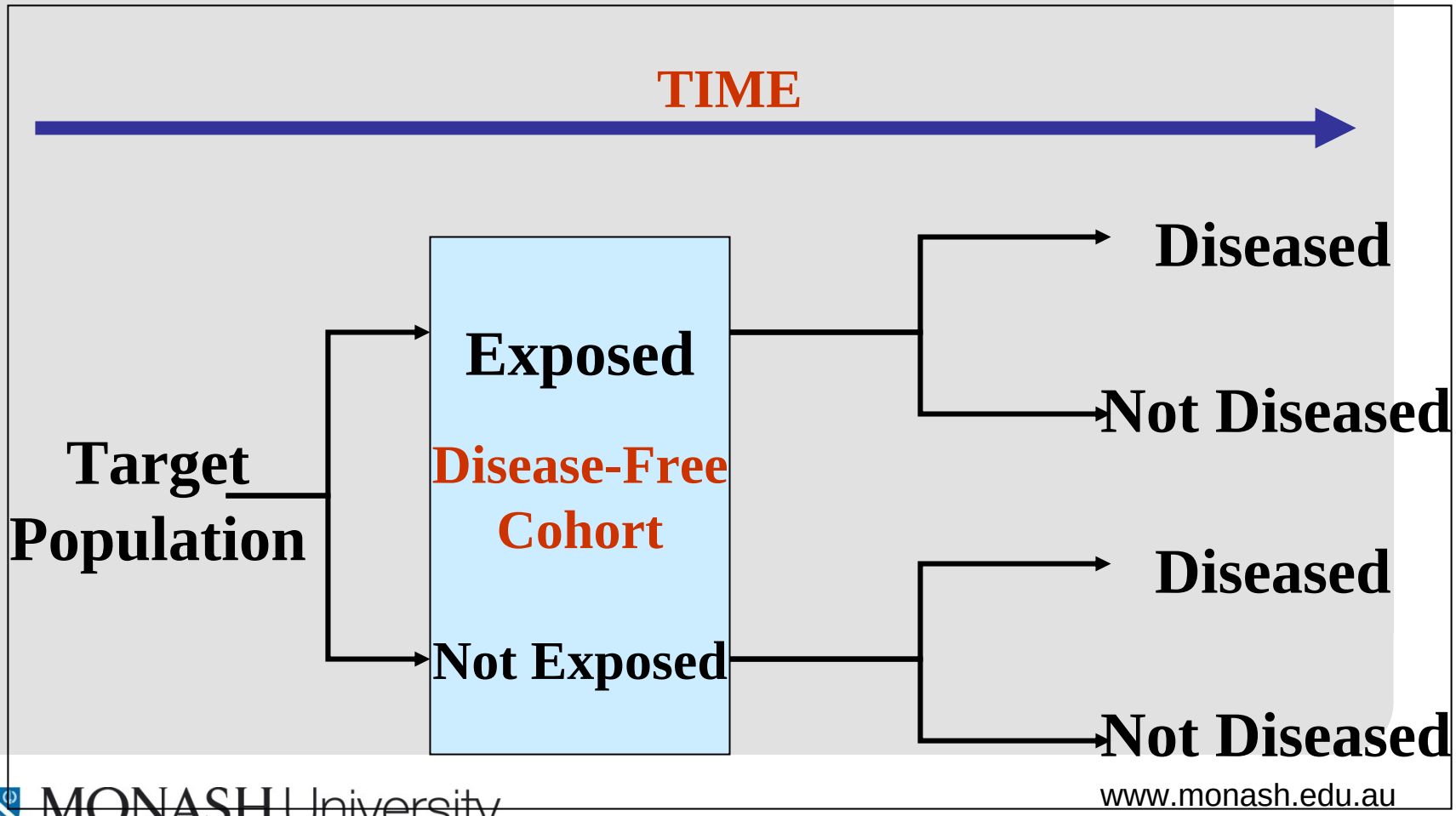
Cohort Studies

- Type of analytical study, so not just descriptive
- Therefore better able to investigate associations between exposure and disease
- Unit of observation and analysis: Individual (not group)
- Time component. Also called follow-up studies, incidence studies, longitudinal studies, or prospective studies

Cohort study design

- At baseline (1st observation point):
  - **Subjects are all disease free**
  - **Exposure is used to classify subjects into exposed or unexposed groups, or different levels of exposure**
- Subjects are followed to document incidence of disease and measure exposure over time (2nd and subsequent observation points)

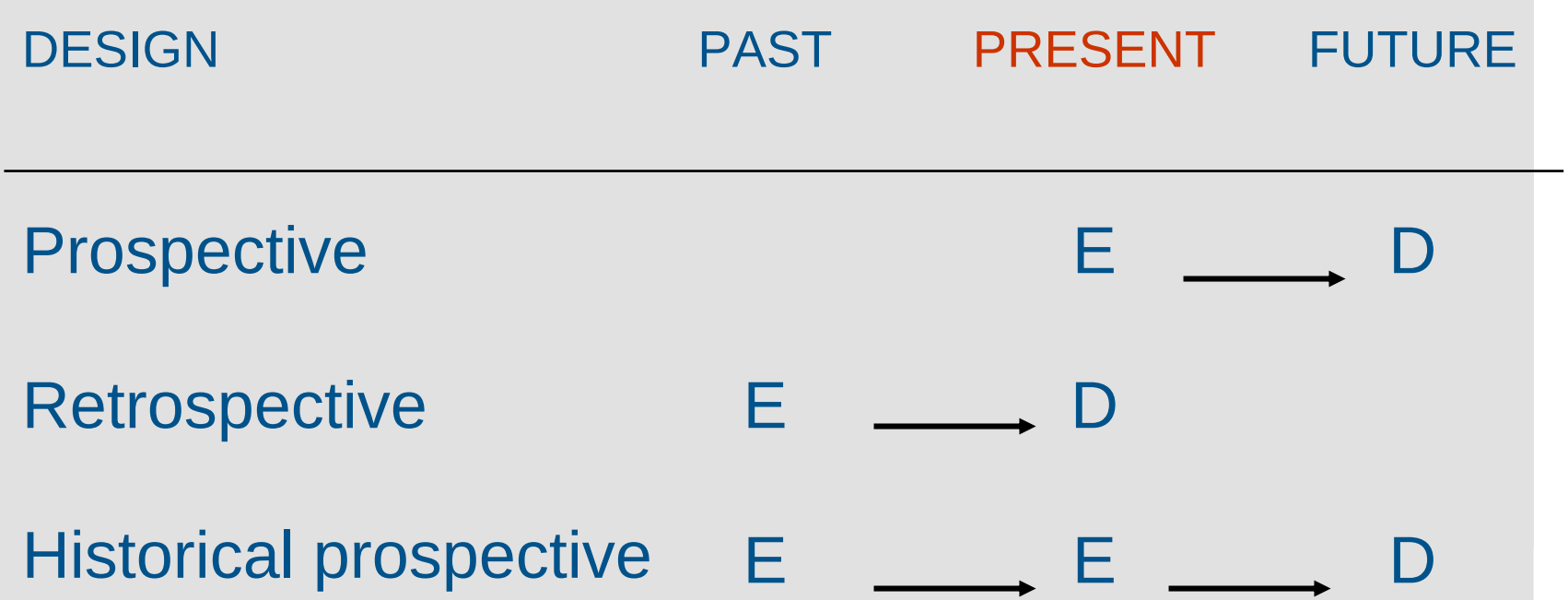
Cohort Study Design



Types of Cohort Studies

- Prospective
  - **Exposure baseline in the present**
  - **Follow-up period: present to future**
- Retrospective:
  - **Exposure baseline in the past**
  - **Follow-up period: past to present**
- Historical prospective:
  - **Exposure baseline in the past**
  - **Follow-up period: past to present to future**

Cohort study types



Historical vs prospective

- Historical much quicker and cheaper to do, as exposure and disease have already occurred
- Reliance on existing records; usually poorer data quality and completeness
- Often no information on potential confounders
- Cancer registry data may be less reliable in the past, eg national data in Australia only since 1983

Assembling a Cohort

- Cohorts may be chosen because they represent:
  - **The general population (where the outcome of interest has a moderately high incidence rate)**
  - **Special exposure groups (e.g., smokers, uranium miners, asbestos workers, pest control operators)**
  - **Special resource groups (e.g., physicians, nurses, military personnel)**
  - **Geographically or facility-defined groups (e.g., Three Mile Island, residents around chernobyl)**

Assembling the Cohort

- May be whole population or a sample
- May be dynamic (people continue to enter the cohort) or closed
- Need a strict definition
- In industrial cohorts, often:
  - minimum employment period
  - too few females for inclusion

The Framingham Study

Since 1948, samples of residents of Framingham, Massachusetts, have been subjects of investigations of risk factors in relation to the occurrence of heart disease and other chronic disease outcomes

The Framingham Study

- Study population consisted of 5,127 men and women between ages 30 and 62 years and were at the time of entry free of cardiovascular disease (1948-1952)

- Cohort was examined for disease outcomes every 2 years and by daily surveillance of hospitalisations at Framingham Hospital

The Framingham Study

- Exposures included:
  - **Smoking**
  - **Alcohol use**
  - **Obesity**
  - **Elevated blood pressure**
  - **Elevated cholesterol levels**
  - **Low levels of physical activity, etc.**

Measuring Disease

- You must determine endpoints in a similar manner for both the exposed and the non-exposed
  - **That is, procedures for disease identification must be the same for the exposed and the non-exposed**
- Define the outcomes of interest (set diagnostic criteria)
  - **If you are looking for multiple outcomes, each must be defined**

Measuring Disease (cont.)

- Mortality may be ascertained from medical records, autopsy records, death certificates, physician records, or next-of-kin
- Hospital records can be scanned for specific types of admissions:
  - **Health records of employers can be monitored, but often poor quality**
  - **Reportable diseases may be ascertained from state registries, eg cancer registries (gold standard)**
- Common ailments that do not usually require medical care may be monitored through self-reports or telephone surveys. Not usually suitable for cancer.

Determining occupational exposure

- Valid means of determining exposure include:
  - **Aim to develop quantitative measures for dose response relationships**
  - **Air monitoring measurements – not for all subjects or jobs or years (may need to impute values)**
  - **Biological tests, eg blood lead levels**
  - **Company records**
  - **Years in job**
  - **For cancer, cumulative exposure most relevant**

Outcome Measures

- Incidence in the exposed
- Incidence in the unexposed
- Relative risk
- Standardised mortality ratio (SMR)
- Standardised incidence ratios (SIR)
- For cancer, incidence is a better measure, as cancer mortality influenced by diagnostic patterns, screening, treatment

INCIDENCE IN COHORT STUDIES

CUMULATIVE INCIDENCE- all cases known to have occurred in the baseline cohort during the duration of the study, divided by the number of individuals enrolled in the study at baseline, per unit time. A risk measure.

INCIDENCE DENSITY-all cases known to have occurred in the total cohort during the duration of the study, divided by the person years of observation contributed by the total cohort per unit time. A rate measure.

In a closed cohort, we can measure either cumulative incidence or incidence density. In an open cohort, we can only measure incidence density.

Incidence Rate (IR)

What is person time?

When we observe a group of individuals for a period of time in order to ascertain the development of a new outcome, such as cancer,….

- The actual time each individual is observed will most likely vary.

Person-Time

Each subject contributes a specific person-time of observation

(days, months, years) to the denominator

Person	Follow-up Time on Study	Person Yrs.
1	<---------->	2
2	<----------D	2
3	<----------WD	1
4	<---------->	3
5	<---------->	2

1995 Jan. 1996 Jan. 1997 Jan. 1998 Jan.

Person-Time

Person	Follow-up Time on Study	Person Yrs.
1	<---------->	2
2	<----------D	2
3	<----------WD	1
4	<---------->	3
5	<---------->	2

1995 Jan. 1996 Jan. 1997 Jan. 1998 Jan.

Number of Cases **1**

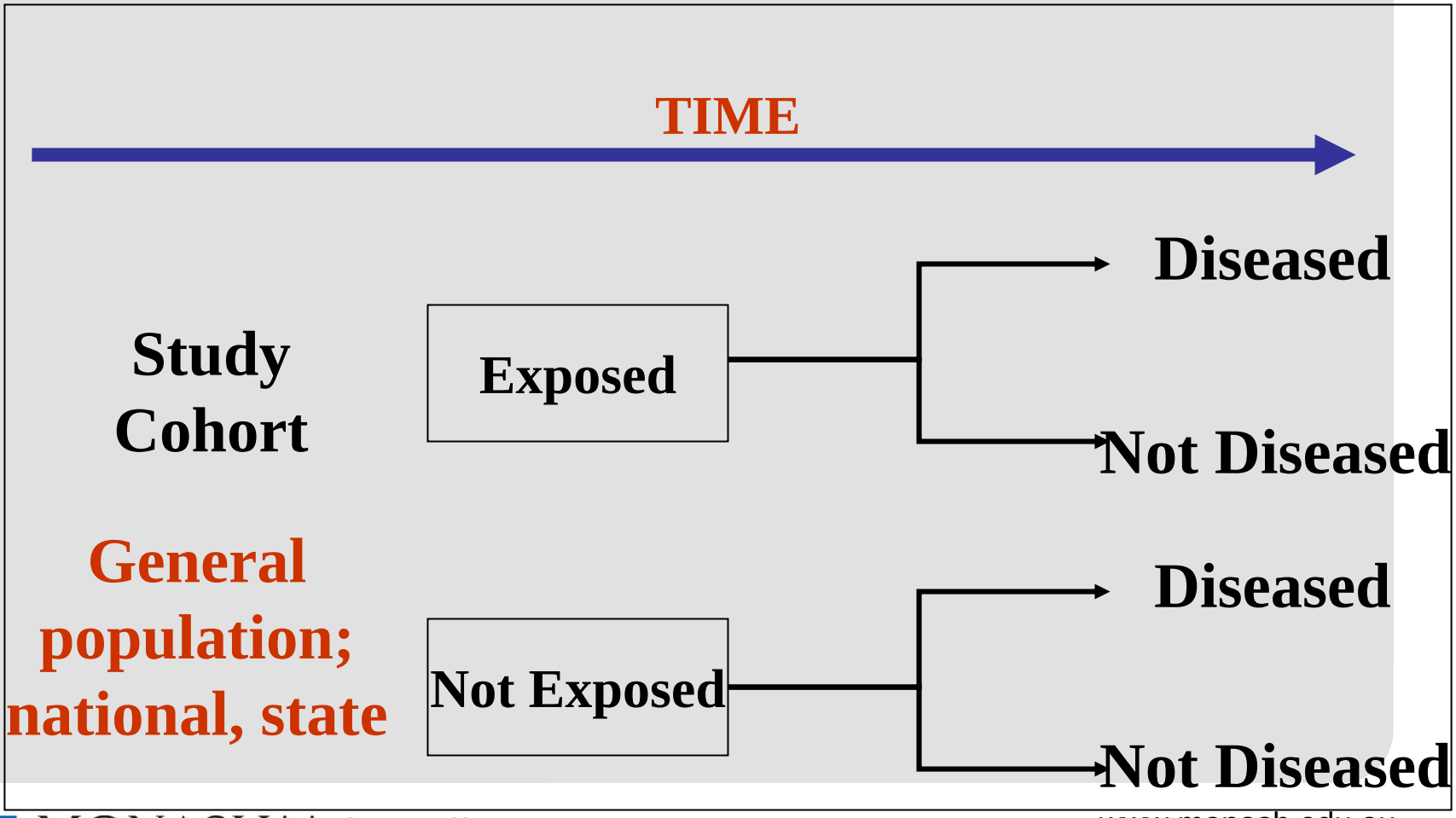
Person Years of Observation: **10**

IR = 1 case / 10 person years of follow-up

External analyses of a cohort study

- Usually link cohort members against a population based cancer registry to find observed cancer cases in the cohort
- Calculate expected numbers for cohort based on population rates and person years at risk
- Calculate a standardised incidence ratio (SIR)
- Often low, in part due to healthy worker effect, whereby entry into industry (and the cohort) influenced by risk factors for cancer
  - eg, respirator use may mean smokers less likely to go into dusty industries
- May be differences in other confounding factors, eg smoking. May not always be able to adjust for these

Linkage cohort study



Internal analyses

- Can help to overcome healthy worker effect
- Compare different exposure groups within the same cohort to develop risk ratios
- Need to be careful about the ‘non-exposed’ part of the cohort, as may differ from the exposed members in important ways, eg SES
- Can overcome this problem by using lowest exposure group as the baseline group

Relative Risk (RR)

- A ratio that measures the risk of disease among the exposed to the risk among the unexposed
- RR Numerator: Incidence in the exposed
- RR Denominator: Incidence in the unexposed

Example: Calculating the Relative Risk

Disease Status

Exposure Status	Cancer (Cases)	No cancer (Controls)	TOTAL
Exposed	112	176	288
Non-exposed	88	224	312

$$\text{Relative Risk} = \frac{A/(A+B)}{C/(C+D)} = \frac{112 / 288}{88 / 312} = 1.38$$

Example: Interpreting the Relative Risk

Relative Risk = 1.38

The risk of developing cancer is 1.38 times higher for an exposed worker compared with a non-exposed (or low exposed) worker.

or

The risk of developing cancer is 38% higher for an exposed worker that a non-exposed (or low exposed) worker.

	RR<1	RR=1	RR>1
Risk comparison between exposed and unexposed	Risk for disease is lower in the exposed than in the unexposed	Risk of disease is equal for exposed and unexposed	Risk for disease is higher in the exposed than in the unexposed
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)

Nested case control study in a cohort

- You may also NEST a case-control study within a cohort study

Example:

- Haematopoietic tumour SIR found to be elevated within the cohort of Australian petroleum workers (Health Watch)
- Cases of leukaemia identified. 5 controls per case from rest of cohort
- More intensive benzene exposure assessment

Nested case cohort study in a cohort

- You may also NEST a case cohort study within a cohort study

Example:

- Shanghai study of cancer incidence in 267,000 female cotton textile workers
- Identify several different series of cancer cases
- Assemble a representative sample of cohort members, which can be used for all case cohort analyses
- Study efficiency, as more detailed exposure assessment only on part of the cohort

Advantages of Cohort Studies

- Temporality: As longitudinal, exposure precedes outcome because the cohort is disease free at baseline
- Efficient for studying rare exposures
- May be used to study multiple disease outcomes
- Measure exposures prospectively
- Allows for calculation of incidence of diseases in exposed and unexposed individuals
- Minimises recall bias, which can be a major problem with case control studies

Disadvantages of Cohort Studies

- Tend to be expensive (large sample size) and time consuming (long follow-up period)
- Loss to follow-up
  - **When multiple outcomes or specific disease incidence is the outcome of interest, bias can be a serious problem**
- Inefficient to study rare diseases

Summary

- Cohort studies considered strongest study design of the observational study designs
- Prospective, so can calculate incidence rate
- Very useful for uncommon exposures
- Less useful for rare outcomes
- Usually better exposure assessment than other designs
- Ca be very time consuming and expensive