Epidemiology 1 - Notes

Introduction to Epidemiology

Epidemiology is a core discipline of public health. **Public health** is the science and art of preventing disease, prolonging life, and promoting health through organised efforts of society.

Epidemiology is the study of the distribution and determinants of disease frequency in human populations and the application of this study to control health problems. It is about measuring disease or other aspects of health, identifying the causes of ill-health and intervening to improve health.

Epidemiology involves:

- Study of populations (groups of people, not individuals).
- Disease frequency
- Disease distribution (patterns of disease according to characteristics of person, place and time).
- Disease determinants (causal and preventative factors)
- Control of health (research to identify determinants, studies testing interventions, surveillance and monitoring).

An **outcome** is a disease, injury of health state of interest (e.g. SIDS). **Exposure** is any factor that might cause/prevent an outcome including unsafe sex, air pollution or vaccination. A **risk factor** is an exposure that is associated with increased/decreased chance of getting an outcome (e.g. smoking is a strong risk factor for lung cancer).

Ethics in epidemiological research include four moral principles: beneficence (do good), non-maleficence (do no harm), respect for autonomy (including privacy and right to make informed decisions) and justice (equity).

Bias involves a lack of validity. **Internal validity** is the extent to which a study's results are valid (i.e. not biased), including selection bias and information bias. **External validity** involves whether the results of a study can be generalised to other populations.

Case Study: Sudden Infant Death Syndrome (SIDS)

SIDS is the unexplained death of an infant (i.e. person of less than 12 months of age). The figure below shows that the number of deaths from SIDS per 100,000 live births was relatively stable in Australia from 1980 to 1990.

Measures of Disease Frequency

In order to monitor the progress towards eliminating existing problems and to identify the emergence of new problems, one needs to be able to quantify the levels of ill health or disease in a population.

There are many types of outcome (disease, injury, death or health status). For health data to be meaningful, diagnostic criteria leading to a **case definition** (decide whether someone has a particular outcome) have to be clear and unambiguous

The type of population affects the types of measures of disease frequency that can be calculated and the types of epidemiological studies that can be done.

A **dynamic population** involves:

- Transient membership based on current status.
- Most geographically defined populations are dynamic (enter and leave).
- Example: current residents of Victoria.

A **fixed (closed) population** involves:

- Permanent membership based on an event.
- Example: everyone who attended a student picnic.

Prevalence is the proportion of people with the outcome at a given time (e.g. what proportion of the Australian population have HIV now?). It is based on 'prevalent' (existing) cases and measures disease/health status. Prevalence is also used to describe frequency of risk factors (e.g. prevalence of smoking) and is an important use for resource planning. It is denoted by:

$$\text{Prevalence} = \frac{Number\ of\ People\ with\ Disease\ at\ Time\ X}{Total\ Number\ of\ People\ in\ Population\ at\ Time\ X}$$

Note that:

- X is usually a specific date.
- Prevalence is a proportion (range is 0 to 1, has no units).
- Suitable for dynamic and fixed population.
- Important for resource planning (incidence rates may decrease, but prevalence can increase, e.g. HIV; thus more resources are required to treat people with outcome).

The **incidence** of disease measures how quickly people are developing the disease. The **incidence proportion** describes the proportion of people who develop the outcome during a specified period:

- Based on new (incident) cases.
- Estimates the probability (risk) of getting the outcome during the period.
- Also called 'incidence risk' or cumulative incidence.
- Requires a fixed population.
- Can be calculated for dynamic populations, but is not practical because it requires knowing when people enter/leave the population.
- Assumes that everyone is followed for the whole period.

- Underestimates the true risk in the presence of competing risks.
- Best used for short periods.

It is denoted by:

$$\text{Incidence Proportion} = \frac{Number\ of\ People\ Who\ Develop\ Outcome\ in\ Specified\ Period}{Number\ of\ People\ At\ Risk\ At\ Start\ of\ Period}$$

Note: use incidence if measuring events ('incidents') onset over period of time (i.e. studying new cases of disease).
Note that when there are competing risks and loss to follow-up, observed IP will underestimate the true risk. Thus, short period of observation is preferred as competing risks and loss to follow-up are minimal.

The **incidence rate** measures the rate (how quickly) people get the outcome. It is based on new cases and is not affected by competing risks and varying lengths of follow-up (can be used to monitor how disease changes over time/determining risk factors). It is denoted by:

$$\text{Incidence Rate} = \frac{Number\ of\ People\ Who\ Develop\ the\ Outcome}{Total\ Person\ Time\ At\ Risk}$$

Note that IR is a rate because time is in the denominator, has units and is suitable for fixed or dynamic populations. **Person time** relates to the time that people are followed.

For a dynamic population living in a geographical area (e.g. Australia), the time that each person is at risk is not usually known. The total person time is hence:

$$Total\ Person\ Time\ at\ Risk = Average\ Population\ Size \times Time\ Population\ Observed$$

Presenting risk is often better than presenting rate for public health communication. Risk (incidence proportion) can be calculated from the rate when it cannot be calculated directly:

$$IP = 1 - e^{-IR \times Time}$$

Here, time is defined as the length of time for which one wishes to estimate the risk (assumes rate is constant over the time period).

Lifetime risk gives the probability that a person will develop a disease (or die) by a specific age, chosen to represent the expected lifetime. It is calculated from current age-specific rates of the outcome of interest.

Crude rates describe the overall incidence or death rate in a population without taking any other features of the population into account (e.g. different age structures).

Age-specific rates describes separate rates for different age groups. **Sex-specific rates** describe separate rates for other groups, e.g. men and women.