# **Lecture 1: Epidemiologic Measures Part I**

# **Learning Objectives**

By the end of this session, participants should be able to:

1. Understand the need for denominators in epidemiology and know how person-time at risk is estimated in follow-up studies.
2. Calculate different measures of disease frequency (prevalence, incidence rate, risk and odds) and appreciate the distinction between them.

# **1. Introduction: definition of a case**

Epidemiologic research is based on the ability to quantify the occurrence of disease in populations. This requires a clear definition of what is meant by a **case**, i.e. the individual in a population who has the disease, health disorder, or suffers the event (e.g. death) of interest.

It is rarely easy to define what is meant by a ‘case', even for a well-known disease. The epidemiologic definition of a case is not necessarily the same as the clinical definition and, in some circumstances, epidemiologists may have to rely on screening tests that are less invasive and cheaper than the diagnostic tests used by clinicians. Cases may be identified by using registries and notifications, abstracts of clinical records, surveys of the general population, etc.

In this module we will concentrate on the situation where individuals can acquire the disease of interest, and therefore become cases, only one time (e.g. HIV infection).

Once a ‘case definition’ has been decided, the number of cases in a population can be determined. To make informed decisions epidemiologists use two other pieces of contextual information about the number of cases: 1) the size of the population from which the cases originated, and 2) the time period during which the cases were counted.

# **2. Measures of disease frequency**

There are two major types of measures of disease frequency: **prevalence** and **incidence**.

## **2.1. Prevalence**

The **prevalence** is the number of cases of disease in a population at one point in time, as a proportion of the total number of persons in that population.

Thus, this measure can be thought of as the frequency of the disease in a population at a point in time and that is why it is sometimes referred to as the **point prevalence**.[[1]](#footnote-1) And as such, that point in time should be specified with the prevalence figure. It is often referred to as a ‘snapshot’- a measure from a particular point in time.

Prevalence is a measure of disease occurrence which can be obtained from surveys (see lecture on Descriptive, Ecological and Cross-sectional studies). It estimates the burden of disease in a population. Such information is useful to public health planners and administrators who wish to determine the allocation of health care resources in a particular community, and need to know what services are required to respond to needs in the population.

A related measure is the prevalence odds, which has the same numerator as the prevalence:

For the prevalence odds, the sum of the numerator and the denominator is equal to the total number of people in the population.

## **2.2. Incidence**

The number of cases of a disease present in a population at a point in time depends not only on the frequency with which new cases occur and are diagnosed, but also on the average duration of the disease (i.e. time to either recovery or death). As a consequence, prevalence figures may vary across populations (or vary within the same population over time), solely because of variations in duration of the disease.

In contrast to prevalence, incidence quantifies the number of new cases of disease that develop in a population of individuals at risk during a specified time interval. Because incidence concerns new cases, it is the preferred epidemiologic measure for establishing causal relationships. Three distinct measures of incidence may be calculated:

### 2.2.1. Incidence risk

**Incidence** **risk** is the proportion of persons in a population comprised of people who are initially free of disease, who subsequently develop the disease within a specified time interval:

This measure of incidence can be interpreted as the probability that an individual will develop a disease during a specified period.

The likelihood of becoming a ‘case’ will increase with the duration of follow-up, and thus the time period to which it relates must always be clearly specified (e.g. 2-year risk, 5-year risk etc).

A second measure of incidence is the **incidence** **odds**, which is the number of new cases divided by the number of non-cases.

For the incidence odds, the sum of the numerator and the denominator is equal to the total number of people at risk at the start of the time period.

### 2.2.2. Incidence Rate

The incidence risk assumes that the entire population at risk at the beginning of the study period has been followed up for the entire specified time period. However, it is often the case that participants are enrolled a study over a period of time, not all at once. Furthermore, during the course of the study some participants may be lost to follow up (e.g. because they move or die or decline to continue participating). In these instances, the length of follow-up will not be uniform for all participants. And for individuals who develop the disease, no account is taken of *when* disease onset occurred during the follow-up period.

To account for these varying time periods of follow-up, we use a denominator which is equivalent to the sum of each individual's time at risk, i.e. the sum of the time that each person remained under observation and at risk of becoming a case.

The figure below shows how **person-time at risk** is calculated using a hypothetical group of 9 persons.



The incidence rate is:

Like incidence risk, the numerator of the incidence rate is the number of new cases in the population. The denominator, however, is now the sum of each individual's time at risk. In the above hypothetical example, the incidence rate is:

In presenting an incidence rate, it is essential to specify the time period to which it refers and the relevant time units - that is, whether the rate represents the number of cases per person-day, person-month, person-year, etc. For ease of presentation, it is possible to say that the rate was 9.4 new cases per 100 person-years, or 94 cases per 1000 person-years.

It is sometimes necessary to estimate rates from vital statistics data when measures of each individual’s person-time at risk are not available. In this case, the population at the mid-point of the calendar period of interest, multiplied by the length of the period (in suitable units of time - usually years) may be taken as the estimate of the person-time at risk.

For example:

|  |  |
| --- | --- |
| Cancer deaths in population A, 2004-2008: | 150 |
| Population in mid-2006: | 22,554 |
| Estimated total person-years at risk: | 5 (years) x 22,554  = 112,770 person-years |
| Mean annual cancer death rate, 2004-2008: | = 150 deaths/112,700 person-years  = 0.0013 deaths per person-year  = 13 deaths per 10,000 person-years |

(Note that a death (or mortality) rate is just an incidence rate with a case being defined as a death.)

The above definitions are now widely accepted, but the terms risk and rate are used interchangeably in much of the (especially older) literature.

## **2.3. The relation between prevalence, incidence and duration of disease**

Non-diseased population

Diseased population

Incidence rate

Recovery rate

Entry

e.g.

birth

Exit e.g. death

Prevalence reduces if a) people are cured or b) if people die or c) if people leave your study population for other reasons. So, if a disease is typically fatal, then introducing an intervention that prevents death (but doesn’t cure), means that prevalence is likely to increase (i.e. more people are living with the disease in the population), depending on the duration of the disease.

A population in which the numbers of people with and without the disease remain stable is known as a steady-state population. In such (theoretical) circumstances, the point prevalence of disease is approximately equal to the product of the incidence rate and the mean duration of disease (i.e. length of time from diagnosis to recovery or death), providing that the prevalence is less than about 10%.

Although true steady-state conditions are never met in practice, the above relationship provides a useful `rule of thumb' for making rough estimates of the prevalence in real populations when no dramatic changes in the incidence or duration of disease have occurred.

# **References**

Webb P and Bain C. *Essential Epidemiology: An introduction for Students and Health Professional*s. Chapter 2 and 5. Second Edition. Cambridge University Press. 2011.

Bailey L, Vardulaki K, Langham J and Chandramohan D, *Introduction to Epidemiology* Chapter 2 and 3*.* Open University Press, 2005 (Understanding Public Health, Series editors: Nick Black and Rosalind Raine).

Hennekens CH & Buring JE, *Epidemiology in Medicine*, Chapters 4 and 12. Little, Brown and Company, 1987.

1. **Period prevalence** is a variation which represents the number of persons who were a case at any time during a specified (short) period divided by the total number of persons in that population. [↑](#footnote-ref-1)