SECTION 5: General Epidemiology

Chapter **19**

Epidemiologic Approach to a Disease

Epidemiology is a relative new field in science. This gained popularity from 1940 to 1970s. This period is known as the 'golden era' of epidemiologic development.

EPIDEMIOLOGY

The study of distributions and determinants of health-related states or events in specified populations and the applications of this study to control health problems.
                                                          JM Last 1988

It is basically the study of how diseases occur in populations and why.

Epidemiology is about information: Information needed for health planning, supervision and evaluation of health promotion and disease control activities. Key components of the data needed may be approached through a series of questions.

Who? Who is affected? Refers to age sex, social class, ethinicity, occupation, heredity and social factors (these are the person factors).

Where? Where did it happen? In relation to place of residence, geographical distribution and place of exposure (place factors).

When? When did it happen? In terms of months, season or years (time factors).

What? What is the condition, its clinical manifestation and diagnosis?

How? How did the disease occur? In relation to the interplay of specific agent, vector and source of infection, susceptible groups and other contributing factors.

Why? Why did it occur? In terms of reasons for the disease outbreak.

Why now? The most important question. What action to be taken as the result of the information gained.

Basic Tenets of Epidemiology

Oriented Towards Groups than Individuals
When clinical observations are made on individual patients, epidemiological observations are made on group of people

(population). Therefore, statistical methods are needed for analysis of epidemiological data.

All Findings must Relate to a Defined Population
It is not enough merely to list down people with a particular disease. Population at risk of developing that disease also has to be enumerated.

Enumerating 'Population at risk' is an important step in epidemiology.

Table 19.1 may convince a person that between 1999 and 2004, the Cholera situation has worsened in the city 'X' (Fictious data).

When the same data is analyzed after accounting for the population of the city during the same period (Table 19.2), it can be noted that Cholera incidence has hardly changed during the period 1999–2004. This shows the importance

Table 19.1 Cholera cases in city 'X' 1999–2004

Year	Cholera cases
1999	505
2000	600
2001	700
2002	800
2003	1,000
2004	1,200

Table 19.2 Annual incidence of cholera in city 'X' (1999–2004)

Year	Population	Cholera cases	Annual incidence/1,000 population
1999	3,500,000	505	0.14
2000	43,00,00	600	0.13
2001	49,00,000	700	0.14
2002	53,00,000	800	0.15
2003	62,00,000	1,000	0.16
2004	72,00,000	1,200	0.16

of taking into account not just the numerator, but also comparing it with a valid denominator. Epidemiologists are people who are in search of a denominator.

Conclusions are Based on Comparisons

Comparing the rates of disease frequency among the exposed (e.g. smokers) and non-exposed (e.g. non-smokers) is the classic epidemiological method.

Description of Events by Time, Place and Person

Here the occurrence of an event is described in terms of when it occurred, where it occurred and who were affected. The classic sequence an epidemiological study is shown in Box 19.1.

Uses of Epidemiology

- Description of health status of population.
- Describing natural history of disease.
- Identify causative factors.
- Evaluation of clinical signs, symptoms and decision analysis.
- Evaluation of interventions or programs.
- Community diagnosis.
- Syndrome identification.

Box 19.1 Epidemiological sequence

- Observation
- Enumerating cases or events
- Relating cases or events to the population at risk
- Making comparisons
- Developing hypothesis
- Testing the hypothesis
- Make scientific inferences
- Conduct experimental studies
- Intervention and evaluation.

COMMUNITY DIAGNOSIS

Community diagnosis involves identifying and quantifying the health problems of a community in terms of mortality and morbidity rates and ratios.

This differs from clinical epidemiology or clinical medicine in that stress is given to the population or community than an individual.

Community diagnosis often helps one to understand the "submerged portion of the iceberg of disease".

Advantages

- Helps in the identification of 'at risk groups' in the community.
- Formulate priorities in disease control and prevention based on the findings.
- The mortality and morbidity ratios may be used for evaluation of health services at a later date.
- Provides us with an in-depth knowledge on the natural history of disease.
- Community diagnosis is not applicable to disease alone, but also examines the social, cultural and environmental characteristics of the community.

Role of Clinical Epidemiology

Clinical epidemiology is the application of epidemiological principles and methods to the practice of clinical medicine.

Clinical decision should be based on sound scientific (evidence-based medicine) observation.

Areas of Clinical Epidemiology

- Definitions of normality and abnormality.
- Validity and accuracy of diagnostic and screening tests.
- Natural history and prognosis of disease.
- Effectiveness of therapies.
- Prevention of disease in clinical practice.

Measuring Magnitude of a Disease

BASIC TOOLS FOR MEASURING MAGNITUDE OF A DISEASE

The basic tools for measuring magnitude of a disease in epidemiology are:
- Rates
- Ratio
- Proportion.

Rate

Rate measures the "occurrence of a particular event in a population during a given time period".

This indicates the change in some event taking place in a population over a period of time.

$$Death\ rate = \frac{No.\ of\ deaths\ in\ a\ year}{Mid\text{-}year\ population} \times 1{,}000$$

Others are specific and standardized rates.

Essential Components
As seen in the above formula rate comprises of:
- Numerator
- Denominator
- Time specification
- Multiplier.
  Rate is expressed as 1,000, 10,000, 100,000, etc. Examples are crude death rate, birth rate, prevalence rate, etc.

Ratio

Ratio is relation in size between two random quantities.

Essential Features
- Numerator not a component of denominator
- Numerator and denominator may involve an interval of time or may be instantaneous in time.

Sex ratio =

$$\frac{No.\ of\ females\ in\ a\ given\ population\ in\ a\ year}{No.\ of\ males\ in\ the\ given\ population\ in\ that\ year}$$

Other examples: Doctor-population ratio, child-women ratio.

Proportion

Proportion indicates the relation in magnitude.

Essential Features
Numerator is always included in denominator.

Case fatality rate

$$= \frac{No.\ of\ deaths\ due\ to\ disease}{Total\ no.\ of\ cases} \times 100$$

Proportional mortality rate

$$= \frac{No.\ of\ deaths\ from\ a\ specific\ disease}{Total\ no.\ of\ deaths} \times 100$$

MEASURES OF MORTALITY

Measures of mortality are listed in Box 20.1. Figure 20.1 shows the number of cancer deaths up to the year 2000 in Country 'X'. Clearly the number of people dying from cancer is increasing significantly through the year 2000, but from this graph, one cannot say that the risk of dying from cancer is increasing, because the only data available is number of deaths (numerators). Denominator (Population at risk) is not specified.

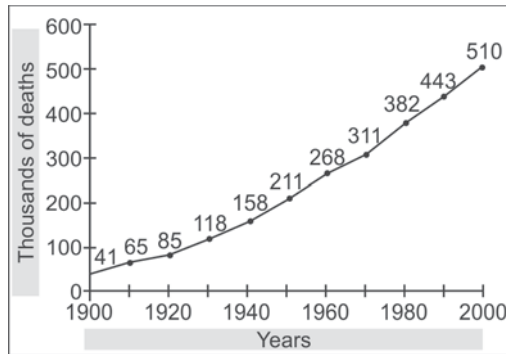| **Box 20.1** Measures of mortality |
|---|
| **1. Crude death rate** = $\dfrac{No.\ of\ deaths\ during\ the\ year\ from\ all\ causes}{No.\ of\ persons\ in\ the\ population\ at\ mid\text{-}year} \times \boldsymbol{1000}$ |
| **2. Specific death rate (lung cancer)** = $\dfrac{No.\ of\ deaths\ from\ lung\ cancer\ per\ year}{No.\ of\ persons\ in\ the\ population\ at\ mid\text{-}year} \times \boldsymbol{1000}$ |
| **3. Case fatality rate** = $\dfrac{Total\ no.\ of\ deaths\ due\ to\ a\ particular\ disease}{Total\ no.\ of\ cases\ due\ to\ the\ same\ disease} \times \boldsymbol{100}$ |
| **4. Proportionate mortality** = $\dfrac{No.\ of\ deaths\ from\ a\ particular\ disease}{Total\ deaths\ from\ all\ causes} \times \boldsymbol{100}$ |

Fig. 20.1 Measures of mortality

Mortality Rates

To express mortality in quantitative terms, there should be a denominator from which subjects have the potential to enter the group represented by the numerator.

Types

Some important mortality rates are given below.

- *Crude death rate*

$$= \frac{No.\ of\ deaths\ during\ the\ year\ from\ all\ causes}{No.\ of\ persons\ in\ the\ population\ at\ mid\text{-}year} \times 1,000$$

This is also termed as 'annual mortality rate'. Another modification of crude death rate is specific death rates.

- *Specific death rate:* This is done on the basis of age, sex, race, disease condition, etc.

Specific Death Rates

- Age-specific death rates for children under 10 years (per 1,000 population):

$$= \frac{\begin{array}{c}No.\ of\ deaths\ from\ all\ causes\ in\\ 1\ year\ in\ children < 10\ years\end{array}}{\begin{array}{c}No.\ of\ children\ in\ the\ population\\ < 10\ years\ at\ mid\text{-}year\end{array}} \times 1,000$$

- Specific death rate from lung cancer:

$$= \frac{No.\ of\ death\ from\ lung\ cancer\ per\ year}{No.\ of\ children\ in\ the\ population\ mid\text{-}year} \times 1,000$$

Time specification is important in mortality rates.

Disadvantage of Crude Death Rate

Crude death rates lack comparability for population with difference.

Advantages of Age-specific Death Rates

- Most useful single measure of mortality
- Helps to identify groups 'at risk' for preventive action
- Good civil registration system is a prerequisite for calculation.

Case Fatality Rate

$$= \frac{No.\ of\ deaths\ due\ to\ disease}{Total\ no.\ of\ cases\ of\ that\ particular\ disease} \times 100$$

Case fatality is a measure of severity of a disease.

Importance

- Case fatality measures the killing power of disease.
- Generally applicable to acute infectious diseases. Limited application in chronic diseases.
- CFR is closely related to virulence.
- It can also be used to measure benefits of a new therapy for a disease as condition improves.

How do Mortality and Case Fatality Differ?

In mortality, denominator represents the entire population at risk dying from the disease, but in case fatality denominator is limited to those who already have the disease (Box 20.2).

Proportionate Mortality

Proportionate mortality is another measure of mortality, which is actually not a rate. Proportionate mortality expresses the **number of deaths due to a particular cause (or in a specific age group) per 100 (or 1,000) total deaths.**

- Proportionate mortality from cardiovascular disease =

$$\frac{\begin{array}{c}No.\ of\ deaths\ from\ cardiovascular\\ disease\ in\ a\ given\ year\end{array}}{\begin{array}{c}Total\ no.\ deaths\ from\ all\\ causes\ in\ that\ year\end{array}} \times 100$$

In other words, proportionate mortality rate (PMR) indicates that of all the deaths what proportion were due to cardiovascular disease.

- Under-five proportionate mortality rate =

$$\frac{\begin{array}{c}No.\ of\ deaths\ under\text{-}five\ years\\ of\ age\ in\ a\ given\ year\end{array}}{Total\ deaths\ in\ the\ same\ period} \times 100$$

Advantage: The PMR can give us a quick look at the major causes of death.

Drawback: The PMR cannot tell the 'risk' of dying from a disease. We need a mortality rate for that.

Box 20.2 Comparison of mortality rate and case fatality rate

Assume a population of 100,000 people of whom 20 are sick with disease 'x' and in 1 year, 18 die from disease 'x'.
Mortality rate in that year from disease

$$'x' = \frac{18}{100,000}$$

$= 0.00018$ or 0.018%

Case fatality rate from disease $= 18/2 = 9 \Rightarrow 90\%$

Survival Rate

$$= \frac{\textit{Total no. of patients alive after 5 years}}{\textit{Total no. of patients diagnosed or treated}} \times 100$$

- This is a method of describing prognosis in certain disease conditions
- Yard stick for assessing standards of therapy, e.g. cancer therapy.

Years of Potential Life Lost

In recent years, another mortality index, years of potential life lost (YPLL) has been increasingly used. The concept of YPLL involves estimating the average time a person would have lived had he/she did died prematurely. This measure is used to help quantify social and economic loss owing to premature death, and it has been promoted to emphasize specific causes of death affecting younger age groups.

STANDARDIZATION

Standardization is done for comparing mortality in different populations.

An important use of mortality data is to compare two or more populations or one population in different time periods. Such populations may differ in regard to many characteristics of which age distribution is the most important. Standardization is a technique for comparing mortality between populations, while effectively holding characteristics such as age.

Two Techniques of Standardization

1. Direct standardization
2. Indirect standardization.

Direct Standardization—Steps

1. Select a standard population
   Standard population: This is a population in which numbers in each age and sex group are known. It should be a relevant and larger population than the study population, with ideally a similar age/sex structure.
2. Apply to the standard population, the age-specific rates of the population whose crude death rate is to be adjusted or standardized. This yields an 'expected number of deaths' (events) in the standard population obtained. These are added together for all age groups to give the total expected deaths.
   Rationale for using the standard population: The replacement of the age group frequencies in the study population with those in the standard population gives the rate that would be observed, if the age structure of the study population were the same as that of the standard population. This allows for fairer comparison between study populations with differing age structures.
3. Divide the 'expected' total number of deaths by the total of the standard population, which yields the standardized or adjusted rate (*Tables 20.1 and 20.2*). *For example*: Comparison of death rates of two populations A and B has been done by adjusting with a standard population.

Table 20.1 Steps in direct standardization for A

Age (Years)	Standard population	Death rate in A (per 1,000)	Expected deaths at A's rates
<1	6,000	15	90
1–14	23,000	1.0	23
15–34	41,000	1.0	41
35–54	30,000	4.0	120
55–64	15,000	15	225
>64	35,000	80	2,800
Total	**150,000**	**35.62**	**3,299**

Age adjusted death rate (per 1,000) = 3299 × 1000/1,50000 = 21.993

Table 20.2 Steps in direct standardization for B

Age (Years)	Standard population	Death rate in B (per 1,000)	Expected deaths at B's rates
<1	6,000	20	120
1–14	23,000	1.75	40.25
15–34	41,000	1.0	41
35–54	30,000	5.0	150
55–64	15,000	20	300
>64	35,000	90	3150
Total	**150,000**	**17.4**	**3,801.25**

Age adjusted = 3801.25 × 1,000/150,000 = 25.342 death rate (per 1,000)

Interpretation: Adjusted death rate of population B is more than population A.

Advantages of Direct Standardization

- Two directly standardized rates are calculated using the same standard population and can be compared, and differences tested for statistical significance.
- It is appropriate for trend analysis as rates can easily be compared over time.

Disadvantages of Direct Standardization

When the number of events occurring are small, the estimated rates may be unreliable.

Indirect Standardization

The simplest and most useful form of indirect standardization is the standardized mortality ratio (SMR). This may be used when age specific death rate of the study population is not available.

This is expressed as the total number of deaths that occur in the study group to the number of deaths that would have

Table 20.3 Calculation of standardized mortality ratio (SMR) in Population A

Age (Years)	Population in A	Standard death rate	Observed deaths	Expected deaths in A (at standard rates)
<1	1,000	20	*	20
1–14	3,000	1.75	*	5.25
15–34	6,000	1.0	*	6.0
35–54	13,000	5.0	*	65
55–64	7,000	20	*	140
>64	20,000	90	*	1800
Total	**50,000**	**17.4**	**1781**	**2,036.25**

Expected deaths in population A = 2036.25

Observed deaths in population A = 1781

SMR = Observed deaths/Expected deaths × 100

= 1781/2036.25 = 0.8747 = 87%

Fig. 20.2 International death certificate

been expected to occur, if that study group had experienced the death rates of a standard population.

$$SMR = \frac{Observed\ deaths}{Expected\ deaths} \times 100$$

Calculation of SMR is shown in Table 20.3.

Interpretation: If the ratio had value greater than 100, then the occupation would carry a greater mortality risk than that of a general population. If the ratio had a value less than 100, then the occupation risks of mortality would be less than that of the whole population.

ROLE OF DEATH CERTIFICATES IN MORTALITY MEASUREMENT

Death certificate is the basis of mortality data. Many countries have routine system of collecting mortality data. The WHO has recommended a standardized death certificate for International use as shown in Figure 20.2. A set of questions are added to the basic structure of International death certificate used in India.

The certificate of death is always issued by doctor as an honor/respect to the deceased person without any fee.

In accordance with the Registration of Births and Deaths Act 1969, the registration of deaths is now compulsory through-out India. The doctor must write the registration number in death certificate and register for such information should be maintained with- in clinic/hospital and a copy/information of death must/mandatorily be sent immediately to the Birth and Death Registration Office. It is essential that the cause of death must be documented/determined before lawful disposal of the deceased body by the doctor. The death certificate also provides the exact cause of death for statistical purposes. The cause of death is recorded according to international conventions; the sequence being that adopted by the WHO. Thus, the International medical certificate of the cause of death consists of two parts.

Part I: Records (a) *the immediate cause,* and (b*) the morbid conditions*, if any, giving rise to the immediate cause. Thus (a) must be due to (b). When (b) is due to other causes, it should be mentioned in (c). The basic pathological condition is that on the lower-most line of part 1 and this is the one that is used for statistical purpose. The medical practitioner should use his/her clinical judgment in completing the medical certificate of cause of death.

Measures of Disease Frequency

Measurement of events (disease or health events) is the most important step in epidemiological studies. In other words, quantification is an essential process in epidemiological research. A simple method of doing this is by counting the frequencies or events. The information may not be of much help, if the frequency (numerator) is not related to an appropriate population (denominator). The measure of disease frequency used for quantifying disease depends on what question is being asked.

The terms incidence and prevalence are often used synonymously in medical literature due to lack of awareness regarding basic epidemiological concepts. These definitions should be used with precision in appropriate circumstances.

Let us consider prevalence and incidence in some detail.

PREVALENCE

The number of persons with a disease or an attribute at a specified point of time.

Prevalence per 1000

$$= \frac{\text{No. of cases of a disease present in the population at a specified time}}{\text{No. of persons in the population at that specified time}} \times 1,000$$

Application of Prevalence

Suppose one was interested in finding out how many people living in a village had diabetes (assuming that investigator could perform the standard oral glucose tolerance test on all people living in the village). If 100 out of 1,000 villagers tested were positive for diabetes, is this proportion (10%) an incidence or prevalence figure? This is obviously a prevalence figure as 10% was derived by testing people only at one point in time (a cross-sectional estimate). One has no idea as to when exactly this 10% actually became diabetic.

Many of the epidemiological works yield only prevalence estimates. True incidence figures are hard to find because of the inherent difficulty in estimating incidence.

Factors Determining Prevalence

Diagnostic Methods

Prevalence of diabetes as measured by fasting blood sugar will be different from prevalence as measured by urine sugar.

Population at Risk

It is essential to ensure that only susceptible individuals are included in the denominator, while calculating prevalence (if the prevalence of carcinoma cervix is being estimated the denominator should be only women who belong to a certain age group (25–69).

Severity of the Disease

If the disease being investigated is a fatal one (high case fatality) then the prevalence rate is likely to be low.

Duration of the Disease

Prevalence will be higher for diseases, which lasts for long time (chronic diseases).

Incidence of the Disease

If a disease has a higher incidence and if it lasts for a reasonably long period, the prevalence is likely to be higher.

Migration: Immigration of susceptible people is likely to increase the prevalence, while out-migration of affected cases is likely to decrease the prevalence.

Prevalence Rate

Number of people with disease at a particular time/total number of people at risk.

Although referred to as a rate, prevalence rate is actually a proportion.

Example: Prevalence of HIV positivity in intravenous (IV) drug abusers at a particular time is calculated by:

$$\frac{\text{No. of HIV positive IV drug abusers}}{\text{Total no. of IV drug abusers in population}} \times 1,000$$

Method: Prevalence rates are generally found by prevalence surveys.

Characteristics

- Measures the burden of disease in a population
- Refers specifically to all current cases (old and new).
- It is of two types:
    - Point prevalence
    - Period prevalence.

Significance

Increased prevalence rates signifies:

- Increased duration of an acute illness.
- Reduced mortality by the disease.
- Better supportive medical care.

An increased prevalence rate may merely reflect increased duration of an illness rather than suggesting that members of the population are at greater risk of contracting disease.

Prevalence rates are not generally appropriate for acute illness as it depends on the average duration of the disease.

POINT PREVALENCE (TABLE 21.1)

It is the number of the individuals who have an illness at a specific point in time.

Example: The number of HIV positive people on 1st July 2002 divided by total population on that date.

When the term "prevalence rate" is used it means "point prevalence" unless specified otherwise.

PERIOD PREVALENCE (TABLE 21.1)

It is the number of individuals who have an illness during a specified time period.

Example: The number of people who are HIV positive in 2002 divided by total mid-year population in 2002.

Question	Measure of disease frequency
1. How many people in a given population have the disease at this point of time?	Point prevalence
2. How many people in a given population ever had the disease during a given period of time?	Period prevalence
3. How many people in a given population newly developed the disease during a given period of time?	Incidence

INCIDENCE

The number of new events, i.e. new cases of a disease in a defined population within a specified period of time.

Application of Incidence

Suppose one wants to find out how many people in a given population newly develop diabetes in a certain period of time. Let us assume that all people were screened at the start of the calendar year (January) and 10% of 100 were found to be diabetic. This means that 900 people were nondiabetic or healthy at the start of the year. Let us presume that this 900 population is again screened for diabetes at the start of the next year and 9 people were found positive. This figure of 10% (9/900) (after 1 year) is the 1 year incidence of diabetes in this population.

Estimation of incidence has a longitudinal (follow-up) component in it as compared to prevalence, which has a cross-sectional component in it.

Table 21.1 Point prevalence and period prevalence in studies of asthma

Interview question	Types of measures
Do you currently have asthma?	Point prevalence
Have you had asthma during the last (n)* years?	Period prevalence

*n = Number of years

Types of Incidence

Similar to prevalence, incidence is divided into two types:

1. Cumulative incidence rate.
2. Incidence density.

Cumulative Incidence Rate

Number of new cases of disease occurring over a specified period of time in a population at risk in the beginning of the interval.

Example: Incidence rate of HIV positivity in IV drug abusers is calculated by =

$$\frac{No.\ of\ IV\ drug\ abusers\ newly\ diagnosed\ as\ HIV\ positive\ in\ 2002}{Total\ no.\ of\ IV\ drugs\ abusers\ population\ in\ 2002} \times 1000$$

Incidence can also be expressed as incidence density as shown below.

Incidence Density

Number of new cases of disease occurring over a specified period of time in a population at risk throughout the interval.

The numerator does not differ between the two types of incidence. However, the denominator can differ in incidence density from cumulative incidence because it takes account of (as in the example) drug abusers who left the area during the year, who died and new cases of drug abusers in the population later in the year.

Incidence density requires us to add up the period of time each individual was present in the population and was at risk of becoming a new case of disease. Incidence density characteristically uses as the denominator person-years at risk (*time period can be person months, days, or even hours, depending on the disease process being studied*).

Incidence density =

$$\frac{No.\ of\ people\ who\ get\ a\ disease\ in\ specified\ period\ of\ time}{Sum\ of\ length\ of\ time\ during\ which\ each\ person\ in\ the\ population\ is\ at\ risk\ (total\ person\ time\ follow\mbox{-}up)} \times 1000$$

Incidence rates may be calculated for diseases like acute respiratory infections and diarrheal attacks where disease is expressed in episodes or spells. In this situation incidence may be expressed as Incidence rate = (in spells)

Incidence rates =

$$\frac{\textit{No. of spells of sickness in a defined period}}{\begin{array}{c}\textit{Mean no. of persons exposed}\\\textit{to risk during that period}\end{array}} \times 1000$$

Characteristics of Incidence
- It measures the rate at which new cases are occurring in a population
- Not influenced by duration of disease
- Use generally restricted to acute conditions.

Common Incidence Measures
- Attack rate
- Secondary attack rate
- Hospital admission rate.

Significance
- *Rising incidence rates indicate:*
  - Failure or ineffectiveness of current control program
  - Improved reporting
  - Need of a new control program.
- *Fluctuating incidence rates indicate*: Change in etiology of disease including agent host environmental characteristics.

RELATIONSHIP BETWEEN INCIDENCE AND PREVALENCE (FIG. 21.1)
In a steady state (*i.e. if incidence is not changing, and the population is stable*).

   Prevalence rate = Incidence rate times the duration of disease.

**Prevalence = Incidence × duration (mean) of disease**

<div align="center">or</div>

$$\mathbf{P = I \times D}$$

This formula is valid only, if incidence and duration are stable.
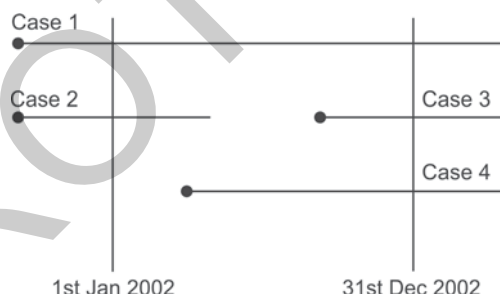
Significance

> *Longer the duration of a disease, the greater its prevalence.*

   Example: Tuberculosis has a higher prevalence rate than incidence because of longer duration of the disease period. Figure 21.1 depicts the relationship between incidence, prevalence, recovery and mortality.
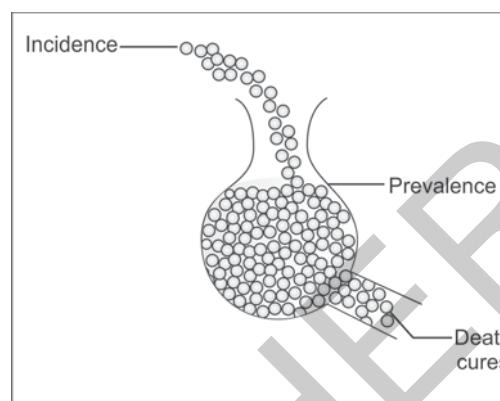
Fig. 21.1 Relationship between incidence and prevalence

Fig. 21.2 Incidence and prevalence

   Figure 21.2 depicts incidence of a disease on 1st January 2002, the rates are detailed below:
- Incidence '0' cases (for the year 2002)
- Point prevalence—2 cases.

On 31st December 2002, the rates are:
- Incidence—2 cases (for the year 2002)
- Point prevalence—3 cases
- Period prevalence—4 cases (for Jan–Dec 2002).

   The incidence and duration of a disease can be derived by the relationship:

> *Incidence = P/D*
> *Duration = P/I*