

MedicalBiostatistics.com

Measures of Morbidity in a Community Prevalence and Incidence

[Measures of morbidity](#)

[Capture-recapture methodology](#)

[Duration of morbidity](#)

[Attack rates](#)

[Disease spectrum](#)

Morbidity is deviation from health. It restricts routine physical, social or mental activities one way or the other. Disease or sickness is the commonly understood manifestation of morbidity but, in wider sense, it can also be unusual values of physiological and biochemical parameters, physical or mental impairment, aberrant behavior such as smoking and sexual malpractices, crimes, and such other conditions that are perceived harmful to the health of people in its holistic sense.

Morbidity greatly varies from time to time, from person to person, and from group to group, and from population to population. How do we measure the magnitude of morbidity in different situations, or how do we assess which person or which population has higher or lower morbidity than the other? Natural measures at individual level are presence or absence of morbidity, and its duration and severity if present. Frequency of occurrence is a legitimate measure for those conditions that can recur. At group and population level, these convert into incidence, prevalence and average duration. Severity at population level is ignored in most situations although it can be measured as percentage of people suffering from different grades of severity. This article provides details of each of these statistical measures although the focus is on incidence and prevalence.

Incidence rate

Incidence rate is the most appropriate measure of frequency of development of morbidity in a group of people over a period of time. Note the word 'development'. In its heart are the new cases among those who do not have that morbidity but are at risk. And period is its essential component. Algebraically,

Incidence rate per unit of time in group A

$$= \frac{\text{Number of new cases arising in Group A over the unit of time}}{\text{Number of persons at risk for developing the outcome}}$$

If 6 new cases of diabetes occur in a group of 90 females of age 50-59 years over a period of two years, the incidence rate per year in this group is $3/90 = 0.033$. Note how this can become an awkward-looking number in some situations. This is the rate per person – in this example per female of age 50-59 years. To make it user-friendly, it is multiplied by an appropriate constant. In this example, you can multiply this by 1000 and say that incidence of diabetes in females of age 50-59 years is 33 per thousand per year. If you so like, you can say 3.3% per year. It may not be proper to say in this case that the incidence is 33% per decade. Precautions are listed later but note that no standards are generally available regarding the unit of time and the unit of base. Uncommon morbidities such as epilepsy can be stated as per million per year and common morbidities such as hypertension in terms of per cent per year. For example, incidence of acute encephalitis in India during 1978-2011 was an average of 0.42 per 100,000 people per year¹. For duration, per year is fairly a standard for most diseases. I will shortly mention a different measure for recurring morbidities.

Incidence is not so easy to obtain for any group of people as it sounds. There are several intricacies and we all should be alive to the implications.

1. For incidence, it is necessary to define onset. Then only you can count. Onset of a morbidity could be in terms of clinical manifestation of signs and symptoms, its detection when screened, unusual values of measurements such as creatinine excretion and cholesterol level, first visit to a health care facility for that condition, etc. Different definitions will give different incidence. While comparing incidence in your study with someone else, make sure that the onset has common definition.
2. Since incidence is related to the duration of follow-up, it is necessary that each person is followed-up for sufficient time for morbidity to appear in at least some of them. For example, for development of cancer, one month of follow-up is ridiculous. For a condition such as this, probably even 10 years may not be enough for general population but may be enough for people with high risk such as heavy smokers. But then the rate you obtain is applicable to heavy smokers only.
3. The formula given earlier assumes that you are able to follow-up everybody for the fixed duration as planned with no dropout or lost to follow-up. This may be difficult in many situations. For varying periods of follow-up, there is a concept of person-time. Details of this are given later.
4. When the follow-up is for two years, as in our diabetes among females example, and the rate is calculated per year, the underlying assumption is that the incidence in the first year is not much different from the rate in the second year. If you are following-up people with liver transplant for development of life-threatening complication, the incidence in first year may be very different from incidence in the second year and in the third year. In this case, rate arrived from 3-year follow-up can not be intrapolated to rate per year.
5. Extrapolation also causes similar problems. A three-month follow-up will not give you correct estimate of one-year incidence. In our diabetes among females example, a two-year follow-up will not tell you the incidence rate per decade.
6. Incidence is a reflection of etiological factors. This is sometimes interpreted as risk of developing the morbidity.
7. Incidence rate may increase if reporting has improved or detection has improved. This might be occurring for many chronic diseases in India. All increases in incidence are not necessarily increase in risk of the disease.

For recurring events such as diarrhoea and angina, attack rate is calculated. This name is generally used for infectious diseases, particularly in case of outbreaks. For attack rate, number of episodes is counted in place of number of persons affected. For example, cumulative attack rate of cholera after earthquake in Haiti in 2010 was 5.1% at the end of first year and 6.1% at the end of second year². Since this is cumulative, the attack rate in second year was only 1%. This incidentally also explains how an incidence rate in the first year after exposure could be very different from second year.

Another related term in the context of infectious diseases is secondary attack rate. This measures the average number of persons infected by the index case. For example, HIV epidemic occurred when the secondary attack rate was more than one and receded when this rate became less than one.

Person-time

When the follow-up period varies from person to person, an epidemiological tool for calculating exposure is person-time. This can easily happen in case of dynamic cohort where people continuously join and leave such as oral contraceptive users who join when they start using the pill and leave when they stop its use. In this example, the calculation will be person-months and not person-years. Varying periods most commonly occur due to dropouts as some people refuse to report back, migrate, become too sick, or die. If one person is exposed or observed for 2 years, second person for 4 years and third person for 1 year, the total person-years are 7. Occurrence of any event of interest during this period can be noted and incidence per person-year or per 100 person-years can be calculated. For example, incidence of depression among non-demented primary care attenders aged 75 years or more in Germany was found 36.8 per 1000 person-years in men and 46.0 in women³.

Person-time tool is often misused. It presumes that incidence in the initial period is the same as in later periods of exposure. For many exposures, this does not hold. In case of complex surgery, the risk of death in first few days or weeks could be very different from subsequent period when the patient stabilizes. In such a situation, person-time can lead to wrong results. A blatant misuse of person-time tool is in calculating smoking exposure in terms of pack-years. Pack-years for a person smoking 40 cigarettes a day for 5 years is the same as for 10 cigarettes a day for 20 years. Both are 200 pack-years. Intensity of smoking as measured by number of cigarettes per day may have different implication than duration for, say, incidence of cardiovascular disease (CVD). This is ignored when pack-years are used as done by Mannan et al.⁴. Incidence of disease per pack-year (or 1000 pack-years) has not been calculated in this paper but if calculated such as 1% risk of biochemical recurrence of cancer per pack-year in cases of radical prostatectomy⁵, this fallacy is obvious.

Prevalence rate

Erroneously called rate, prevalence is the proportion or percentage of people affected at a point of time. This provides a one-time cross-sectional snapshot of the situation in place of film of occurrences. This is calculated by the answer to the question that you have specific morbidity or not at the time of contact. The survey may take weeks, months or years but that is not a consideration in this setup. If somebody is diabetic at the time of contact, he or she is a prevalent case irrespective of when this disease occurred. For example, in a small study in Chandigarh, prevalence of metabolic syndrome was 55% in bipolar disorder patients, 34% in schizophrenic patients and 6% in healthy controls⁶. These groups were consecutive cases and not matched for,

say, age and sex, and since metabolic syndrome is associated with age and sex, these prevalences are not strictly comparable. This illustrates one precaution required in drawing inference from prevalence, in fact almost all such indices. Other issues are as follows:

1. As in the case of incidence, the denominator for prevalence is the number of people at risk for that morbidity. For example, for smoking, children below 12 years can be excluded. Prevalence in specific groups such as in different age-groups and sex should be calculated if these factors can affect prevalence. A case-mix should not happen.
2. For the purpose of comparison between groups, areas, diseases, etc., it is customary to calculate the prevalence *rate* (per cent, per thousand, or per million persons) at a particular point in time. **For uniformity, we suggest that prevalence rate for all diseases should always be calculated per thousand subjects. For a common disease such hypertension in obese, this could be 267 per thousand subjects and for a rare disease such as epilepsy, this could be only 0.17 per thousand persons. When stated with such uniform base, you can see how the rates can be compared across diseases.**
3. Prevalence estimates the probability that a random person in the group has that morbidity. In the context of sensitivity-specificity / predictivity, this is the pre-test probability. If prevalence of abdominal tuberculosis in persons with complaints of constipation, vomiting and pain in abdomen (of long duration) is 22%, this is the chance that persons with these complaints have that disease. After further tests or examination, the (post-test) probability alters depending on the findings.
4. Since prevalence measures how many people are affected, this is used to estimate the load on health care services.
5. As explained a little later, prevalence is greatly dependent on the duration of disease. Duration of coronary diseases is longer among non-smokers compared to smokers. This may give you the result that prevalence of coronary diseases is high among non-smokers. Carefully consider how fallacious this can be.
6. There is another very interesting limitation of prevalence. It is known that many people with coronary disease stop smoking. Thus, in a cross-sectional survey you may find that prevalence of smoking is low in coronary cases! Who will accept this, howsoever correct the information might be. This is called reverse causation.

Although the concept of prevalence is essentially for a point of time, yet there is a concept of period prevalence as well. In some surveys, the question asked is 'are you suffering at this time or in previous one week suffered from disease A'. This includes the cases that have onset more than a week ago but have spilled into the current week as well as those that arose within that week. Answer to such question will give period prevalence. This is obtained mostly for acute conditions.

Prevalence is affected by incidence – higher the incidence, more is the prevalence, but it is more intimately affected by the duration of that morbidity. Some conditions such as diarrhea may last a few days but other conditions such as hypertension and diabetes may require medication for life. Cases with morbidity of long duration tend to accumulate over a period since the exits by way of remission or death is generally small. This increases prevalence (Figure 1). In this figure, incidence is same for both the diseases but prevalence on, say, 7th day is only one for disease with smaller duration and is nine for disease with longer duration.

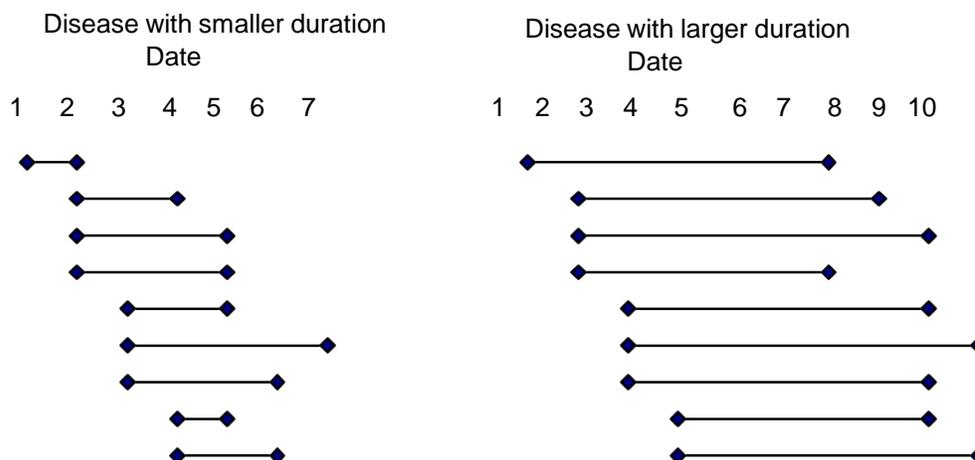


FIGURE 1: Effect of duration of morbidity on prevalence (Source: Ref 7)

Assessment of duration of disease requires that the onset and the termination are sharply defined. I have already mentioned about varying definitions of onset. The termination too can be in terms of disappearance of complaints, ability to resume normal functions of life, discharge from the hospital, negative test, etc. All this can depend on the age and sex for sure but also on nutrition level, economic compulsion, etc.

Now, since prevalence is affected by incidence and duration of morbidity, what is the relationship? Indeed a mathematical relationship exists provided the conditions remain stable during the period of your interest. This is as follows:

Prevalence = Incidence per unit of time * Average duration of disease in the same unit of time

If incidence is per 1000 persons, the prevalence will also be per 1000 persons. If incidence is per year, the duration of disease should also be in terms of years. Thus, if incidence of malaria in an endemic area is 600 per 100,000 population per year, and average duration is 7 days ($=1/52$ years), the prevalence at any point of time is $600/52 = 11.54$ per 100,000. Expect nearly 12 persons sick from malaria at any time in this population of 100,000.

The formula can be inversely used to calculate incidence when prevalence and duration of disease is known. Incidence is difficult to obtain because it requires expensive follow-up study. This formula provides a method to overcome that limitation. There is a big rider, though. This formula is applicable only in stable conditions as already mentioned. There should not be any intervening factor. For example, for diseases such as Alzheimer's, mortality can affect the prevalence. This formula will not be applicable for that disease. Podgor and Leske⁸ have given a procedure for estimating incidence from prevalence of such irreversible diseases.

References

1. Potharaju NR. Incidence rate of acute encephalitis syndrome without specific treatment in India and Nepal. *Indian J Community Med.* 2012 Oct;37(4):240-51. doi: 10.4103/0970-0218.103473.
2. Barzilay EJ, Schaad N, Magloire R, Mung KS, Boncy J, Dahourou GA, Mintz ED, Steenland MW, Vertefeuille JF, Tappero JW. Cholera surveillance during the Haiti epidemic - The First 2 Years. *N Engl J Med.* 2013 Jan 9. [Epub ahead of print]

3. Weyerer S, Eifflaender-Gorfer S, Wiese B, Lupp M, Pentzek M, Bickel H, Bachmann C, Scherer M, Maier W, Riedel-Heller SG. Incidence and predictors of depression in non-demented primary care attenders aged 75 years and older: results from a 3-year follow-up study. Age Ageing. 2013 Jan 11. [Epub ahead of print]
4. Mannan HR, Stevenson CE, Peeters A, Walls HL, McNeil JJ. Age at quitting smoking as a predictor of risk of cardiovascular disease incidence independent of smoking status, time since quitting and pack-years. BMC Res Notes. 2011; 4: 39. Published online 2011 February 15. doi: [10.1186/1756-0500-4-39](https://doi.org/10.1186/1756-0500-4-39)
5. Ngo TC, Lee JJ, Brooks JD, Nolley R, Ferrari M, Presti Jr JC. Smoking and adverse outcomes at radical prostatectomy. Urol Oncol. 2011 Aug 6. [Epub ahead of print]
6. Grover S, Nebhinani N, Chakrabarti S, Avasthi A, Kulhara P, Basu D, Mattoo SK, Malhotra S. Comparative study of prevalence of metabolic syndrome in bipolar disorder and schizophrenia from North India. Nord J Psychiatry. 2013 Jan 7. [Epub ahead of print]
7. Indrayan A. Medical Biostatistics, Third Edition. CRC Press, New York, 2012.
8. Podgor MJ, Leske MC. Estimating incidence from age-specific prevalence for irreversible diseases with differential mortality. Stat Med 1986;5:573-8.