Biomedical and Health Informatics
Vol. 2, No. 1, 2017, pp. 1-4
http://www.aiscience.org/journal/bhi

The Difference between Incidence and Prevalence in Public Health and Epidemiology

Abdulhadi Sale Kumurya[1, *], Sadiya Ashiru Auta[2]

[1]Department of Medical Laboratory Science, Faculty of Allied Health Sciences, Bayero University, Kano, Nigeria
[2]Department of Biochemistry, Faculty of Basic Medical Sciences, Bayero University, Kano, Nigeria

Abstract

Incidence and prevalence are the dominant rates in epidemiology, distinguishable by their respective foci. In other words, incidence is concerned solely with new events, while prevalence considers both new and old cases. These measures are used to produce a number of views on health and disease. Incidence should not be confused with prevalence, which is the proportion of cases in the population at a given time rather than rate of occurrence of new cases. Thus, incidence conveys information about the risk of contracting the disease, whereas prevalence indicates how widespread the disease is. Prevalence is the proportion of the total number of cases to the total population and is more a measure of the burden of the disease on society with no regard to time at risk or when subjects may have been exposed to a possible risk factor. Prevalence can also be measured with respect to a specific subgroup of a population. Incidence is usually more useful than prevalence in understanding the disease etiology: for example, if the incidence rate of a disease in a population increases, then there is a risk factor that promotes the incidence.

Keywords

Epidemiology, Prevalence, Incidence, Rate

Received: December 9, 2016 / Accepted: April 8, 2017 / Published online: June 14, 2017

1. Introduction

This paper is a short communication on the difference between incidence and prevalence. Epidemiologists study the health of population groups within societies or countries. This information helps plan health programmes, to know how one country's health compares with another, and many other things that are useful for all who work in a health setting. Two of the most commonly used terms in epidemiology are prevalence and incidence, and they are often mixed up or used incorrectly.

Prevalence is a frequently used epidemiological measure of how commonly a disease or condition occurs in a population. Prevalence measures how much of some disease or condition there is in a population at a particular point in time. The prevalence is calculated by dividing the number of persons with the disease or condition at a particular time point by the number of individuals examined. For example, in the study above 6139 individuals completed the questionnaire (were examined). Of these 6139 people, 519 currently suffered incontinence and so had the condition at the particular time point of the study. Thus the prevalence of incontinence was 519/6139 = 0.085.

Prevalence is often expressed as a percentage, calculated by multiplying the ratio by 100. The above study expresses prevalence as a percentage, thus the prevalence of incontinence is 8.5% (or rounded is 9%). Another common way of expressing prevalence, particularly if the prevalence is low, is as the number of cases per 100,000 of the population. For example, it is easier to state the result as `66 cases per 100,000 people' than to say the prevalence is 0.00066. Le and Boen [9] provide further examples of the

* Corresponding author
E-mail address: askumurya.med@buk.edu.ng (Kumurya A. S.)

calculation of prevalence.

The incidence of a disease is another epidemiological measure. Incidence measures the rate of occurrence of new cases of a disease or condition. Incidence is calculated as the number of new cases of a disease or condition in a specified time period (usually a year) divided by the size of the population under consideration who are initially disease free. For example, the incidence of meningitis in the UK in 1999 could be calculated by finding the number of new meningitis cases registered during 1999 and dividing that number by the population of the UK. As this incidence rate would be very small again we tend to consider number of cases per 100,000 people.

2. Incidence and Incidence Rate

The incidence of a disease is an epidemiological measure. Incidence is the rate of new (or newly diagnosed) cases of the disease. It is generally reported as the number of new cases occurring within a period of time (e.g., per month, per year). Incidence measures the rate of occurrence of new cases of a disease or condition. Incidence is calculated as the number of new cases of a disease or condition in a specified time period (usually a year) divided by the size of the population under consideration who are initially disease free. For example, the incidence of meningitis in the Nigeria in 2015 could be calculated by finding the number of new meningitis cases registered during 2015 and dividing that number by the population of the Nigeria. As this incidence rate would be very small again we tend to consider number of cases per 100,000 people.

Many research reports talk about incidence, but in strict epidemiological terms there are different concepts that surround the word, including incidence proportion, incidence density and others. However, true incidence is 'the number of instances of illness commencing, or of persons becoming ill' (or dying or being hurt in injuries, or whatever) 'during a given period in a specified population' [8]. When most people use the term they mean the incidence rate, which differs slightly in that it is the rate at which events occur in a population [8]. In other words, incidence usually means something that is measured within a set number of people and in a time period [10, 11]. Incidence can tell us how many new cases of a particular illness have been suffered by a community, or it might tell us how patterns of a condition within a population change over time. Auckland, in New Zealand, often has epidemics of meningococcal disease; with annual incidences of up to 16.9/100,000 people [2]. No one knows why, but by examining ways that the incidence changes over time, and with environmental factors, an understanding of how the disease is spread has been gained. Incidence rates allow comparisons to be made. In the same time frame as the Auckland epidemics were occurring Queensland, in Australia, had an incidence of 2.8/100,000 in one year [14]. Because incidence is expressed as a rate, it does not matter that the population of Queensland, at about 3.5 million people, is larger than the population of Auckland (about 1 million). This information tells health planners that in a year, Auckland suffers, in population terms, a more serious outbreak of meningococcal disease than Queensland. Some care has to be taken when determining incidence as different groups within a community, for example age groups, will need to be considered when doing the calculations.

Incidence is the rate of new (or newly diagnosed) cases of the disease. It is generally reported as the number of new cases occurring within a period of time (e.g., per month, per year). It is more meaningful when the incidence rate is reported as a fraction of the population at risk of developing the disease (e.g., per 100,000 or per million population). Obviously, the accuracy of incidence data depends upon the accuracy of diagnosis and reporting of the disease. In some cases (including ESRD) it may be more appropriate to report the rate of treatment of new cases since these are known, whereas the actual incidence of untreated cases is not [13]. Incidence rates can be further categorized according to different subsets of the population – e.g., by gender, by racial origin, by age group or by diagnostic category.

The incidence rate is the number of new cases per population at risk in a given time period [10]. When the denominator is the sum of the person-time of the at risk population, it is also known as the incidence density rate or person-time incidence rate [8]. In the same example as above, the incidence rate is 14 cases per 1000 person-years, because the incidence proportion (28 per 1,000) is divided by the number of years (two). Using person-time rather than just time handles situations where the amount of observation time differs between people, or when the population at risk varies with time [4]. Use of this measure implies the assumption that the incidence rate is constant over different periods of time, such that for an incidence rate of 14 per 1000 persons-years, 14 cases would be expected for 1000 persons observed for 1 year or 50 persons observed for 20 years [5].

When this assumption is substantially violated, such as in describing survival after diagnosis of metastatic cancer, it may be more useful to present incidence data in a plot of cumulative incidence, over time, taking into account loss to follow-up, using a Kaplan-Meier Plot.

Consider the following example. Say you are looking at a sample population of 225 people, and want to determine the

incidence rate of developing HIV over a 10-year period:

a) At the beginning of the study (t=0) you find 25 cases of existing HIV. These people are not counted as they cannot develop HIV a second time.

b) A follow-up at 5 years (t=5 yrs) finds 20 new cases of HIV.

c) A second follow-up at the end of the study (t=10 yrs) and find 30 new cases.

If you were to measure prevalence you would simply take the total number of cases (25 + 20 + 30 = 75) and divide by your sample population (225). So prevalence would be 75/225 = 0.33 or 33% (by the end of the study). This tells you how widespread HIV is in your sample population, but little about the actual risk of developing HIV for any person over a coming year.

To measure incidence you must take into account how many years each person contributed to the study, and when they developed HIV. When it is not known exactly when a person develops the disease in question, epidemiologists frequently use the actuarial method, and assume it was developed at a half-way point between follow-ups. In this calculation:

a) At 5 yrs you found 20 new cases, so you assume they developed HIV at 2.5 years, thus contributing (20 * 2.5) =50 person-years of disease-free life.

b) At 10 years you found 30 new cases. These people did not have HIV at 5 years, but did at 10, so you assume they were infected at 7.5 years, thus contributing (30 * 7.5) = 225 person-years of disease-free life. That is a total of (225 + 50) = 275 person years so far.

c) You also want to account for the 150 people who never had or developed HIV over the 10-year period, (150 * 10) contributing 1500 person-years of disease-free life.

That is a total of (1500 + 275) = 1775 person-years of life. Now take the 50 new cases of HIV, and divide by 1775 to get 0.028, or 28 cases of HIV per 1000 population, per year. In other words, if you were to follow 1000 people for one year, you would see 28 new cases of HIV. This is a much more accurate measure of risk than prevalence.

3. Prevalence

Prevalence is a frequently used epidemiological measure of how commonly a disease or condition occurs in a population. Prevalence measures how much of some disease or condition there is in a population at a particular point in time. The prevalence is calculated by dividing the number of persons with the disease or condition at a particular time point by the number of individuals examined. For example, in the study above 7234 individuals completed the questionnaire (were examined). Of these 7234 people, 623 currently suffered incontinence and so had the condition at the particular time point of the study. Thus the prevalence of incontinence was 623/7234 = 0.086. Prevalence is often expressed as a percentage, calculated by multiplying the ratio by 100. The above study expresses prevalence as a percentage, thus the prevalence of incontinence is 8.6% (or rounded is 9%). Another common way of expressing prevalence, particularly if the prevalence is low, is as the number of cases per 100,000 of the population. For example, it is easier to state the result as `55 cases per 100,000 people' than to say the prevalence is 0.00055. Le and Boen [9] provide further examples of the calculation of prevalence.

Prevalence (or to be more correct, prevalence proportion and sometimes point prevalence) gives a figure for a factor at a single point in time [7]. We may want to know how many children are overweight in a country. By measuring the height and weight of every child of a particular age in a sample of children and calculating their body mass index, we will have some idea of how many children, at that point in time, in that sample of children, (not necessarily new cases) are overweight. This is the prevalence. The important words are ‘at that point in time’ because prevalence can tell us only what is happening at a certain point. A recent Scottish study showed that the prevalence of obesity in a group of children aged from three to four years was 12.8 per cent at the time when the data were collected [1]. When to use prevalence or incidence depends on what you want to know. Administrators in paediatric hospitals know that the prevalence of respiratory infections is highest in the winter months, so plan staffing levels accordingly. Hence, children’s hospitals in Britain make sure few of their staff are on holidays during January when the prevalence of respiratory disease is highest [1].

Prevalence is the actual number of cases alive; with the disease either during a period of time (period prevalence) or at a particular date in time (point prevalence). Period prevalence provides the better measure of the disease load since it includes all new cases and all deaths between two dates, whereas point prevalence only counts those alive on a particular date [6]. Prevalence is also most meaningfully reported as the number of cases as a fraction of the total population at risk and can be further categorized according to different subsets of the population.

4. Incidence to Prevalence

The relationship between incidence and prevalence depends greatly on the natural history of the disease state being reported. In the case of an influenza epidemic, the incidence may be high but not contribute too much growth of

prevalence because of the high, spontaneous rate of disease resolution. In the case of a disease that has a low (or zero) cure rate, but where maintenance treatment permits sustained survival, and then incidence contributes to continuous growth of prevalence. In such cases, the limitation on prevalence growth is the mortality which occurs in the population. Obviously, prevalence will continue to grow until mortality equals or exceeds the incidence rate [12].

Incidence should not be confused with prevalence, which is the proportion of cases in the population at a given time rather than rate of occurrence of new cases. Thus, incidence conveys information about the risk of contracting the disease, whereas prevalence indicates how widespread the disease is. Prevalence is the proportion of the total number of cases to the total population and is more a measure of the burden of the disease on society with no regard to time at risk or when subjects may have been exposed to a possible risk factor. Prevalence can also be measured with respect to a specific subgroup of a population. Incidence is usually more useful than prevalence in understanding the disease etiology: for example, if the incidence rate of a disease in a population increases, then there is a risk factor that promotes the incidence.

For example, consider a disease that takes a long time to cure and was widespread in 2002 but dissipated in 2003. This disease will have both high incidence and high prevalence in 2002, but in 2003 it will have a low incidence yet will continue to have a high prevalence (because it takes a long time to cure, so the fraction of individuals that are affected remains high). In contrast, a disease that has a short duration may have a low prevalence and a high incidence. When the incidence is approximately constant for the duration of the disease, prevalence is approximately the product of disease incidence and average disease duration, so *prevalence = incidence × duration*. The importance of this equation is in the relation between prevalence and incidence; for example, when the incidence increases, then the prevalence must also increase. Note that this relation does not hold for age-specific prevalence and incidence, where the relation becomes more complicated [3].

5. Choice of Incidence or Prevalence

Generally, when studying reasons for disease, one is most likely to use the incidence rate. This is due to the fact that the incidence rate is not affected by changes in the method of care or by case-fatality, making differences between populations simpler to explain. For research on the weight of disease over short periods (e.g. headache), the incidence rate is also most likely to be used, as point prevalence would underestimate the issue. This is due to the fact that the point prevalence value would overlook the survivors and deaths.

When considering the weight of chronic diseases, even those that are rare (e.g. ALS), one is most likely to use the prevalence rate. For studies of disease risk factors, health activities, and even causes of disease, the prevalence rate is the favoured gauge. Many epidemiologists often see the prevalence rate as a substandard measure in comparison to the incidence rate; however, this is not the case, as each measure has both pros and cons, making their respective value dependent on the issue under investigation.

References

[1] Armstrong, J. and Reilly, J. J. (2002). Child health information team. *Lancet*. 359: 2003-2004.

[2] Baker, M. (2000). Household crowding a major risk factor for epidemic meningococcal disease in Auckland children. *Pediatric Infectious Disease Journal.* 19(10): 983-990.

[3] Brinks, R. (2011). "A new method for deriving incidence rates from prevalence data and its application to dementia in Germany".

[4] Coggon, D., Rose, G. and Barker, D. J. (1997). *"Quantifying diseases in populations". Epidemiology for the Uninitiated (4th ed.). BMJ.*

[5] Dunn, O. J. and Clark, V. A. (2009). *Basic statistics: a primer for the biomedical sciences (4th ed.). Hoboken, N. J.: John Wiley & Sons. pp. 3–5.*

[6] Greenberg, R. S., Daniels, S. R., Flanders, W. D., Eley, J. W. and Boring, J. R. (2005). Epidemiologic Measures, Chapter 2. In: Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring JR, eds. Medical Epidemiology. 4th ed. New York: McGraw-Hill.

[7] Jekel, J. F. (2001). *Epidemiology, Biostatistics, and Preventive Medicine (2nd edition)*. Philadelphia, WB Saunders.

[8] Last, J. M. (2001). *A Dictionary of Epidemiology (4th edition)*. Oxford, Oxford University Press.

[9] Le, C. T. and Boen, J. R. (1995). Health and numbers: basic biostatistical methods, John Wiley, Chichester.

[10] Rothman, K. J. (2002). *Epidemiology: An introduction.* Oxford, Oxford University Press.

[11] sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/.../EP713_DiseaseFrequency7.htmlJun 8, 2016

[12] Walters, R. W. and Kier, K. L. (2012). The Application of Statistical Analysis in the Biomedical Sciences, Chapter 8. In: Kier KL, Malone PM, Stanovich JE, eds. Drug Information: A Guide for Pharmacists. 4th ed. New York: McGraw-Hill.

[13] Waning, B. and Montagne, M. (2001). Medical Surveillance and Outbreaks of Disease, Chapter 2. In: Waning B, Montagne M, eds. Pharmacoepidemiology: Principles and Practice. New York: McGraw-Hill.

[14] Ward, J. (2000). Enhanced surveillance for meningococcal disease in Queensland in 1999. *Communicable Diseases Intelligence.* 24(11): 332-225.