

Frequency measures of epidemiological studies

Abstract

This article explained the measures of frequency and association that are used in observational epidemiological data analysis. The observational studies include cohort, case-control and cross-sectional. In epidemiology, most of the variables are nominal with only two categories like exposed or unexposed, male or female, case or control, so ratios, rates, and proportions are used in the analysis of these types of dichotomous variables. Different fictional data from different studies were used to calculate the incidence rate, relative risk, mortality rate, odds ratio and prevalence of diseases. The odds ratio and relative risk are called measures of association simply because they quantify the relationship between exposure and outcome. Incidence rate, relative risk, and mortality rate were calculated in cohort studies; the odds ratio was determined in a case-control study while prevalence was calculated in a cross-sectional study. The appropriate measure to be used depends on the type of the research.

Keywords: frequency measure, observational studies, *human immuno virus*, epidemiology

Volume 5 Issue 1 - 2017

Ilker Etikan, Sulaiman Abubakar, Rukayya Alkassim

Department of Biostatistics, Nicosia-TRNC, Cyprus

Correspondence: Ilker Etikan, Near East University Faculty of Medicine Department of Biostatistics, Nicosia-TRNC, Cyprus, Email ietikan@gmail.com

Received: December 13, 2016 | **Published:** February 01, 2017

Abbreviations: RR, relative risk; CI, confidence interval; OR, odds ratio

Introduction

People are suffering from some diseases like cancer, HIV/AIDS, diabetes, hypertension, heart disease, malaria, sickle cell anemia, among others. Polio is also one of the diseases suffered by children in some developing countries. A lot of health-related problems have been bedeviling people all over the world. How to measure the diseases, determining their causes and plan the appropriate means of controlling the diseases as well as their occurrences are very important issues address by 'Epidemiology'. Measures of frequency and association are very useful for that purpose and they are regarded as the fundamental of descriptive epidemiology.

Epidemiology is defined by¹ as "the study of the distribution and determinant of health-related events in a specified population and the impact of this study to control of health related problems". Any variable or factor that can affect the frequency of the occurrence of disease in a population is referred to as 'determinant'.²

Epidemiology is very important field that is uses by government, health organizations, among others, in determining the important aspects of human conditions in a particular population. Such aspects include nationality, morbidity and mortality and they are described by rates, ratios and proportions. The main concern of epidemiology is to measure health, discover what bring about the disease and intervene to cure the disease and overcome its causes.³ The role of epidemiology is beyond just a disease but the improvement of health, the control of the disease and devising structure for the health-related problems analysis. Epidemiology is made up of two study designs.

Experimental study design and observational study design are the two basic study designs in epidemiology. In experimental studies, intervention is made by a researcher to modify reality and then observe what will happen, while in observational studies, a researcher notices what occurs but does not make any modification.⁴ Randomized controlled trials and Quasi-experimental design are the types of experimental study designs while cohort study, case-control study as well as cross-sectional study are the three most common types of observational study designs.

Data analysis is very crucial in epidemiological research as it assists in forming and structuring the findings from different sources of data collection and it also helps to keep human bias away from conclusion with the aid of appropriate statistical treatment.⁵ Since some of the variables typically used in observational studies are dichotomous then the measures of frequency and association are used in making the analysis of data to determine the occurrence of disease and/or to measure the association/relationship between exposure and outcome.

This article focuses on some measures of frequency and association calculated for cohort study, case-control study and cross-sectional study.

Material and methods

Fictive data will be used to calculate some measures of frequency and association. Those measures include incidence rate, relative risk, mortality rates, odds ratio and prevalence. Incidence rate, relative risk, and mortality rate will be used in a cohort study to calculate the different rates. Odds ratio will be determined in a case-control study while for cross-sectional study, prevalence will be calculated. Fictional data from the study of cholera cases among Yobe state civilian will be used to calculate the incidence rate while cases for lung cancer fictive data in Jigawa state will be used to analyze the relative risk as well as the odds ratio. For mortality rate, Fictional data of maternal deaths of Kano state will be used to calculate the rate. The fictive data for the survey of patients at a sexually transmitted disease clinic in Kano state will be used to calculate the prevalence of condom used during the specified period.

Observational studies

In this *type of studies*, a researcher observes and systematically gathers relevant information, but does not attempt to modify the subjects being observed. Unlike experimental studies where a researcher intervenes to alter something (e.g., gives a drug to treatment group) and then observes what will occur, no intervention is made by the researcher in an observational study. Examples of observational studies include a survey of smoking habits among adolescents, the study of breast cancer among women aged between 25 and 60, and a study of maladaptive behaviors among high school students.

Observational studies are carried-out when a researcher cannot perform an experiment, when the experiment is not accepted or when the study is not experimental in nature. It is also carried-out when the primary aim of the researcher is to get descriptive information. Cohort study, case-control study, and cross-sectional study are the three most common observational studies (Figure 1).

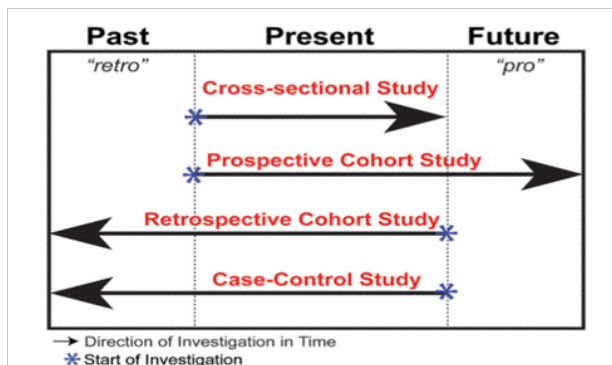


Figure 1 Cross-sectional studies are the three most common observational studies.⁸

Analysis of observational studies using frequency measures

The ratios, proportions, and rates are used in epidemiology to describe the birth, disease and death. The birth rate, mortality rate and the prevalence or incidence rate of a disease can be calculated using the data derived from the observational studies.

It is vital to consider the concept of 'confidence interval' because of error of random sampling in observational studies and the outcome achieved may differ from the reality, because of chance. Confidence interval will be calculated to assess or evaluate the possible impact of this sampling error. The most commonly used confidence intervals in health-related research are 95% intervals. For Relative Risk (RR), the null value or 'no-effect' is 1.0. 1.0 RR indicates that the two groups being compared do not differ. If both ends of the confidence interval are less than 1.0, then it indicates an inverse relationship between exposure and outcome; similarly a positive relationship exists if both ends of the CI are greater than 1.0. However, if the CI includes the null value, i.e. the upper limit is greater than 1.0 and the lower limit is less than 1.0, then a researcher may not disclose the likelihood that the real RR is 1.0, and thus the relationship do not exist between exposure and outcome.³

Cohort study: incidence rate, relative risk and mortality rate

Data from cohort study can be evaluated and/or analyzed using incidence rate, relative risk and mortality rate. Mortality rate is regarded as a descriptive frequency measure while incidence rate and relative risk as measures of comparative effect.⁶ Cohort study analysis used the ratio of the rate of disease in the exposed group compared with the rate in the unexposed group.

Incidence rate

In epidemiology, incidence simply means the occurrence of new cases of disease, for example, new cases of Ebola disease, Lassa fever, or injury in a population during a specified period. The incidence of a particular disease measures how quickly or frequently the disease

of interest is been developed by people. Unlike prevalence, incidence considers only new cases, and it has a unit. In order to measure the incidence of a disease, a cohort study should be conducted. The study will include participants who are at risk of developing the disease of interest. Then they should be followed to determine those that truly developed the disease. Incidence rate is one of the approaches of measuring the frequency of disease in a population. Therefore, the incidence rate of a disease measures the frequency of the disease occurrence in a population over a specified period. Incidence rates are subject to change over time, from disease to health, therefore the period of the cohort need to be specified.

$$\text{Incidence rate} = \frac{\text{Number of new cases during a given time period}}{\text{Total number of people in the population}} \times 10^n$$

Example

The cholera new cases among the Yobe state of Nigerian civilians population is 545 while the Nigerian civilian population was estimated to be 828,262. The cholera incidence rate for the Nigerian civilian population will be calculated using these data.

$$\text{Incidence rate} = \frac{545}{828,262} \times 10^5$$

$$= 0.000658 \times 100,000 = 65.8 \text{ per } 100,000$$

The above example shows that 545 represent the new cases of diseases which were diagnosed during the specified period of the study while 828,262 is the population at risk. This implies that persons who are involved in the 828,262 should be able to develop the disease, which is been described during the period covered.

Relative risk (RR)

Relative risk also called risk ratio, is a measure of relationship which compares the rates of disease in two groups. The rate for the group of primary interest, for example, treatment group, is divided by the rate for a comparison group, for instance, control group. Relative measures are used to detect the frequency of the likelihood of experiencing a particular health outcome for a person who is exposed to something than a person who is not exposed. The measures give a clue about the strength of relationship between the exposure and the outcome, but do not express anything about the definite number of occurrence of disease in either group.

$$\text{Incidence rate} = \frac{\frac{\text{Number of exposed deceased patients}}{\text{Number of exposed persons}}}{\frac{\text{Number of non-exposed deceased patients}}{\text{Number of non-exposed persons}}}$$

Example

2.2% was reported as the risk of lung cancer among smokers while 0.7% was the risk for non-smokers in Jigawa state. The relative risk of lung cancer for the two groups of people (smokers versus non-smokers) is calculated as:

$$\text{Relative risk (or risk ratio)} = 2.2\% / 0.8\% = 2.75$$

The lung cancer risk in smokers is 2.75 of the risk of non-smokers. In other words, the result shows that smokers are more likely to develop lung cancer than the non-smokers.

Mortality rate

This is one of the frequency measures that measure the occurrences of deaths in a given population.⁷ Defined mortality rate as a measure of the frequency of death occurring in a defined population during a specified time interval. There is a need to know the size of the population in which the deaths occur and the total number of deaths during a given period in order to calculate the mortality rate.

$$\text{Mortality rate} = \frac{\text{Number of deaths in a period}}{\text{Number of persons - years}} \times 10^n$$

Example

The following table will be used to calculate the mortality rate for maternal deaths in Kano state.

From the above table, the mortality rate for the entire population will be calculated as:

$$\begin{aligned} \frac{\text{Number of Maternal deaths}}{\text{Population}} \times 10^n \\ = \frac{53}{1597} \times 10^5 \end{aligned}$$

Therefore 3318.7 maternal-related deaths were determined for the given population, and this is calculated per 100,000 population.

Case-control study: the odds ratio (OR)

Odds Ratio is a great measure of association used in a case-control study.⁸ The odds ratio is a relative measure of risk used to determine the likelihood of developing the outcome for a person who is exposed to the factor as compared to that who is not exposed. It is used to evaluate the risk of a particular disease (or outcome) if certain factor (or exposure) is present. When events are rare, risk and the odds are very similar, and it is very easier to interpret relative risks than odds ratio. Thus in many situations, researchers will be able to interpret odds ratios by assuming or pretending that they are relative risks (Table 1).

Table 1 Mortality rate for maternal deaths in Kano state

Age group	Maternal deaths	Population
15-25	21	782
26 – 35	17	540
36 – 45	14	231
≥ 46	1	44
Total	53	1597

Case-control study results can be presented in a form of table (2×2) as follows (Table 2):

Table 2 Case-control study results

	Cases	Controls	Total
Exposed	a	b	a + b
Unexposed	c	d	c + d
Total	a + c	b + d	a + b + c + d

The odds ratio is calculated as:

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

a = number of subjects with both exposure of interest and disease

b = number of subjects with exposure of interest, but without disease

c = number of subjects without exposure of interest, but with the disease

d = number of subjects without both exposures of interest and disease

a + c = total number of subjects with disease (cases)

b + d = total number of persons without disease (controls)

The OR is calculated as a comparative effect measure, and therefore, it is used to determine the strength of relationship that exists between exposure and outcome.

Example

Table 3 shows the totals for females and males. The table will be used to determine the odds ratio.

$$\text{Odds Ratio} = \frac{41 \times 1,156}{1,240 \times 16} = 2.4$$

Table 3 can also be used to calculate the risk ratio: to calculate the risk ratio of pellagra for females versus males, the risk of illness among females and also among males has to be calculated.

$$\text{Risk of illness among females} = \frac{a}{a+b} = \frac{41}{1,281} = 0.032$$

$$\text{Risk of illness among males} = \frac{c}{c+d} = \frac{16}{1,172} = 0.014$$

Therefore, the risks of illness among females and males are 0.032 or 3.2%, and 0.014 or 1.4% respectively. Females are the group of principal interest while males are the comparison group.

$$\text{Risk Ratio} = \frac{3.2\%}{1.4\%} = 0.014$$

The lung cancer risk in females is 2.3 times higher than that of males. The results indicated that the odds ratio of 2.4 and the risk ratio of 2.3 are close to each other. That is one of the interesting features of the odds ratio: when the outcome is not common, the odds ratio provides an appropriate approximation of the relative risk.

Table 3 Number of cases for lung cancer by sex

	Yes	No	Total
Female	a = 41	b = 1,240	1281
Male	c = 16	d = 1,156	1172

Cross-sectional studies: prevalence

A cross-sectional study has to be representative of the whole population. Therefore, appropriate probability sampling technique need to be used in determining the sample size which will represent

the population. In random sampling, each element/participant has an equal chance of being participated in the study through the use of a procedure of random selection.⁹ For instance, a study of the prevalence of hypertension among men aged 40-70 years in Kano city should comprise a random sample of all men aged 40-70 years in that city. Thus, hypertension male patients in Kano city who fall within the stated range (40-70) have an equal probability of being participated in the research.

Prevalence is one of the three important measures that form the fundamental of descriptive epidemiology.¹ The prevalence of a disease is used to determine the proportion of a population that really has the disease of interest at a specific period. It is mainly the outcome measure obtained from a cross-sectional study that measures the occurrence of existing disease. it is influenced by the incidence and the duration of the condition.⁴ Prevalence has no unit.

$$\text{Risk Ratio} = \frac{\text{Number of people with cases at a given point in time}}{\text{Total number of people in the population}} \times 10^n$$

Example

120 of 360 patients interviewed as the reported use of a condom at least once during the three months before the interview in a survey of patients at a sexually transmitted disease clinic in Kano state.

The prevalence of condom use in this population over the last three months is calculated as:

$$\frac{120}{360} \times 10^2 = 0.333 = 33.3\%$$

Therefore, the prevalence of condom use during the given period study was 33.3% in this population of patients.^{10,11}

Conclusion

The cohort, case-control as well as the cross-sectional studies are the three mostly used observational study designs. Measures of frequency and association were used to determine the rates, ratio and proportion for the three observational study designs. Those measures include incidence rate, relative risk, mortality rate, odds ratio and

prevalence. The choice of the measure to be applied in a particular study depends on the type of the study. The confidence interval is a crucial aspect that is suggested to be included in the further review when making an analysis of data especially in determining the relative risk and odds ratio. All the data used in this article is fictive.

Acknowledgement

None.

Conflict of interest

None.

References

1. Last JM. Dictionary of epidemiology. 4th edn. USA: Nw york: Oxford University Press. 2001;61p.
2. Epidemiology: some basic concepts and definitions.
3. Webb P, Bain C. Essential epidemiology: An introduction for students and health professionals. UK: Cambridge University Press. 2011;461pp.
4. http://www.med.uottawa.ca/SIM/data/Study_Designs_e.htm
5. <http://www.dissertationindia.com/blog/importance-of-data-analysis-in-research>
6. Rensing M, Blettner M, Klug SJ. Data analysis of epidemiological studies. *Deutsches Arzteblatt Int.* 2010;107(11):187–192.
7. CDC. Public health surveillance. Lesson 5, Principles of Epidemiology. 2nd ed. USA: Atlanta: Self-Study Course 3030–G, Field test version 4. 1992.
8. Etikan I, Alkassim R, Abubakar S. Meta-Analysis of Using Both Cohort and Case Control Study. *Biom Biostat Int J.* 2016;3(3):00068.
9. Etikan I, Musa SA, Alkassim RS. Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics.* 2016;5(1):1–4.
10. http://www.med.uottawa.ca/SIM/data/Epidemiology_rates_e.htm
11. Song JW, Chung KC. Observational Studies: Cohort and Case-Control Studies. *Plastic and Reconstructive Surgery.* 2010;126(6):2234–2242.