Measures of Frequency in Epidemiology

Matúš Maciak - mmatthew@matfyz.cz

(Exam topic, 23^{rd} of January, 2008)

What is **Epidemiology**? If we consider the definition of Cole (1979), we could say that Epidemiology is a study process in order to investigate the **occurrence** of some illness (illness under the study problem). There are different approaches used in order to provide some research (to answer the study question) - diagnostic research, etiognostic research and pronostic research. Each of them requires some other methodology. Let us focus on that word "occurence" for a while. If we are interested in the occurence of some physical or mathematical event - frequency (f) variable is used to describe it. The world's common definition of frequency says that **frequency is a measure of the number of occurrences of a repeating event per unit of time**. However, mostly it refers to a periodical and reqular changes during the time period. In case of epidemiologic research we will have to use to a little bit different meaning of it. Therefore, i will focuss now some of the most commonly used measures of frequency in epidemiology and the ways of their application in a real epidemiologic study process.

Population at risk

Firstly, for epidemiological purposes the is no reason to talk about some periodical and even regular occurrences of the event. Moreover, the occurence of cases we are studing (the occurrence of disease cases) has to be always related to something. That "something" we call the **population at risk**.

Generaly we can say, that population in risk is the group of people, healthy or sick, who will be taken as cases if they have the disease being study - it is the population of individuals that gives rise to new cases! The right choice of the population at risk is a crutial task for each epidemiologic research.

Frequency measures

The main task in each epidemiologic research is to quantify the occurence of illness and to give or describe relations between the occurrence of disease and characteristics of individuals. There are three main tools used to do that.

1. **Incidence rate** - a measure of the instantaneous force of disease occurrence.
2. **Cumulative incidence** - a proportion of individuals going to the group of diseased from the group of nondiseased, given a certain time period.
3. **Prevalence** - a proportion of people who have the disease at a specific time.

In next parts I will give a small description and comparison of relations between those three measures.

Incidence rate

The incidence rate is **the rate at which new cases occur in a population at risk DURING A SPECIFIED PERIOD OF TIME**! In this case it is insufficient to take the incidence rate just as a simple proportion of population that is affected - one has to give a relation with respect to time period, otherwise it is not usable quantity (just consider the ratio of deaths within a certain population). The definition of incidence rate is given by

$$\mathit{Incidence\ rate} = \frac{\#\left\{\text{disease onsets}\right\}}{\sum_{i=1}^{\text{size of population}} T_i} \in [0, \infty], \qquad (\heartsuit)$$

where each T_i is the period of time for each individual while being observed as a member of population at risk. Sometimes it is refered to as **incidence density** or **force of morbidity**.

- **problem 1:** - the number of onsets for one individual ("first occurrence")
- **problem 2:** - repeated time measures in case of multiple inclussion (more occurrences)
- **problem 3:** - interpretability ONLY with respect to a specific time ("person-time")
- **problem 4:** - type of population (dynamic vs. fixed) ⇒ different time measures!

Fixed population vs. dynamic one

It is important to fully understand the different approach to incidence rate with respect to these two population. While the first one is specified by members at the begining (no additional entrance is allowed) the second one is specified by some characteristics mostly (is open to additional entrance). In case of fixed population the measure of time period for each individual begins at the begining of the study (all population members are there) and it goes on till the first occurence of the case (which does not have to happen) when it stops. In such a case, denominator in (♡) is given as an area under the rate curve. In the case of dynamic population one has to record not only occurrences of cases but also time of the entrance of each new member of the population at risk. The number of individuals at the begining of study has nothing to do with number of individual in population or with the denominator in (♡).

If the number of people entering the population is balanced with the number of individuals exiting it, the papulation is said to be in a ***steady state****.*

Cumulative incidence ratio

Cumulative incidence is defined as **the proportion of a fixed population that becomes diseased in a stated period of time.** In other words, if there is a specific probability (risk) given that an individual will develop the disease in a specified period of time, than the cumulative incidence is the **measure of average risk**. This type of measurement is used more often than the incidence rate given above. The reason for that is a simplier interpretability. **However, one has to be aware that the cumulative incidence is given conditionaly on the EXISTENCE OF NO OTHER RISK OF DEATH!** Cumulative incidence measure is given by

$$CM_t = \frac{P_0 - P_t}{P_0} \in [0, 1], \qquad (\diamondsuit)$$

which is the number of individuals exiting the population during the time period of t in nominator and the initial size of the population in denominator. This refers to fixed population only! Based on these two expression given in (♡) and (◇) one can easily get a relation between the incidence rate and the cumulative incidence ratio as

$$\mathit{Incidence\ rate}(t) = \frac{-\Delta P}{P_t \Delta t}, \qquad \text{and} \qquad CM_t = 1 - \exp\left\{-\int_0^t \mathit{Incidence\ rate}(t)\mathrm{d}t\right\}.$$

- **problem 1:** - competing risks of death - just a hypothetical measure!
- **problem 2:** - not usefull for long periods of time - related to problem 1!

Prevalence
Firstly, there is an important difference between incidence measures stated before and the prevalence measure. While incidence measures focus on events, **the prevalence measure focuses on the disease status.** In other words, **the prevalence is the proportion of a population that is affected by the disease under the study question at the given point in time.** We can describe prevalence as:

1. prevalence is a measure for occurence of status-type illness
2. depends on the risk of geting the illness
3. depends on the risk of remaining in the ill status group (cure or die)
4. approximately equal to product of incidence and mean duration of illness

Beside the original population consider a subpopulation - population of all ill individuals. Sometimes this subpopulation is refered to as **prevalence pool.**. What is interesting is, that each individual has a chance to exit the prevalence pool not only by cause of death but also by recovering from the disease. This is connected to the **steady state** we have mentioned before. If the number of individuals entering the prevalence pool is equal (with respect to some time period) to individuals exiting it, the prevalence pool is said to be **balanced**.

Let N be the total number of people in population and let P stands for the size of the prevalence pool. Then the number of individuals who enter the prevalence pool during some time interval $[t_1, t_2]$ is given by

$$N_{\text{enter}} = \textit{Incidence rate} \times (t_2 - t_1) \times (N - P),$$

and the number of individuals who exit the prevalence pool is defined as

$$N_{\text{exit}} = \textit{Incidence rate}_{\text{exit}} \times (t_2 - t_1) \times P,$$

where the $\textit{Incidence rate}_{\text{exit}}$ is the incidence rate with respect to population in the prevalence pool. As the reciprocal of an incidence rate in a steady state is equal to the mean duration before the occurrence of illness, the reciprocal of incidence rate with respect to population in the prevalence pool is the mean duration of the illness ($\bar{D}$).

For small prevalence values it holds

$$\textit{Prevalence} \doteq I\bar{D}.$$

Prevalence odds
The prevalence odds ratio is defined as **the ratio between the prevalence quantity and the quantity 1 minus the prevalence**. In other words, it is the chance of having the disease with respect to the chance of having not the disease.