

# Measures of Disease Frequency: Prevalence and Incidence

Marlies Noordzij<sup>a</sup> Friedo W. Dekker<sup>b</sup> Carmine Zoccali<sup>c</sup> Kitty J. Jager<sup>a</sup>

<sup>a</sup>ERA-EDTA Registry, Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, and <sup>b</sup>Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands; <sup>c</sup>CNR-IBIM, Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Renal and Transplantation Unit, Ospedali Riuniti, Reggio Calabria, Italy

## Key Words

Prevalence · Incidence · Risk · Incidence rate · Dynamic population · Cohort

## Abstract

To describe how often a disease or another health event occurs in a population, different measures of disease frequency can be used. The prevalence reflects the number of existing cases of a disease. In contrast to the prevalence, the incidence reflects the number of new cases of disease and can be reported as a risk or as an incidence rate. Prevalence and incidence are used for different purposes and to answer different research questions. In this article, we discuss the different measures of disease frequency and we explain when to apply which measure.

Copyright © 2010 S. Karger AG, Basel

## Introduction

The focus of epidemiology is to study the occurrence and determinants of disease. Measuring the frequency of a disease or other health outcome in a population and identifying how the disease frequency may differ over time or among subgroups are important steps in discovering potential causes of a disease and determining effec-

tive methods for prevention and care. There are several measures of disease frequency in common use and depending on the research question and the available data, one should choose which measure is appropriate. In this paper, we discuss the most important measures of disease frequency, i.e. the prevalence, the risk, and the incidence rate. In addition, we explain which of these measures can be used in specific situations and study populations, such as in dynamic populations and in cohort studies.

## Study Populations

To decide which measure of disease frequency one should use, it is helpful to have some understanding of the 2 main types of study populations. The first type is the dynamic population, also referred to as a dynamic cohort or an open cohort. For a dynamic population we assume that, on average, all subjects who drop out of the study for any reason will be replaced by new subjects. Subjects can leave or enter the study at any moment. Examples of dynamic populations are the general population, patients entering and leaving a hospital department, and armies.

In contrast to a dynamic population, a cohort is population with a fixed membership. This means that once the cohort is defined and follow-up begins, no one can be

added. Because a proportion of the initially included subjects die, are lost to follow-up, or develop the disease of interest during follow-up, the composition of the cohort changes and it becomes smaller over time. Cohort studies are commonly applied in clinical research and well-known examples of them are the large prospective cohort studies such as the Framingham Heart Study [1]. Also randomized controlled trials are a special type of cohort study.

### Prevalence

The prevalence represents existing cases of a disease and can be seen as a measure of disease status; it is the proportion of people in a population having a disease:

$$\text{Prevalence} = \frac{\text{Number of subjects having the disease at a time point}}{\text{Total number of subjects in the population}}$$

The prevalence is often useful as it reflects the burden of a disease in a certain population. This is not limited to burden in terms of monetary costs; it also reflects burden in terms of life expectancy, morbidity, quality of life, or other indicators. Knowledge of the burden of disease can help decision makers to determine where investments in health care should be targeted. For instance, the prevalent number of end-stage renal disease (ESRD) patients predicts the need for dialysis facilities and the related costs.

As an example, in a study that was published in 2007, Zelmer [2] assessed the economic burden of ESRD in Canada. A prevalence-based approach was used to estimate direct health care costs associated with ESRD. The author reported that by the end of the year 2000, there were an estimated 24,921 Canadians living with ESRD and the total direct costs of ESRD in that year were 1,273 million dollar.

### Incidence

While the prevalence represents the existing cases of a disease, the incidence reflects the number of new cases of disease within a certain period and can be expressed as a risk or an incidence rate.

### Risk

The risk is the probability that a subject within a population will develop a given disease, or other health outcome, over a specified follow-up period. It can be calcu-

lated by dividing the number of subjects developing the disease over a certain period by the total number of subjects followed over that period:

$$\text{Risk} = \frac{\text{Number of subjects developing the disease over a time period}}{\text{Total number of subjects followed over that time period}}$$

This risk can be interpreted as an estimation of the risk of disease in an individual subject. However, to interpret a risk appropriately, it is necessary to know the time period to which it applies. Without the definition of a time period, a risk is a meaningless value.

This can be illustrated by means of the following example: Ojo et al. [3] studied patient and allograft outcomes among African-American kidney transplant recipients with ESRD as a result of sickle cell nephropathy when compared to all other causes of ESRD. They identified 22,647 patients who received a renal transplant between 1984 and 1996. Of them, 82 had sickle cell nephropathy. The risk of acute rejection in the 1st year after transplantation was 43.9% in the recipients with sickle cell nephropathy compared to 41.9% in those with other causes of ESRD.

One can imagine that without reporting the time period to which they applied, the rejection rates of 43.9 and 41.9% could not have been interpreted. A risk of rejection of 43.9% within 1 week would have been extremely high, while the same risk over a period of 50 years would have been surprisingly low.

For the calculation of a risk, a few assumptions need to be made. First, because the risk reflects new cases of disease, all subjects should be free of this disease at the start of the follow-up period. Second, all subjects should be followed over the total period of time during which the risk is measured. This second requirement can lead to a few problems, especially in studies with a relatively long follow-up period. The longer the follow-up period is, the higher is the chance subjects will become lost to follow-up. In addition, subjects can drop out of the study because of causes that 'compete' with the outcome of interest. For example, if one aims to study death of cardiovascular causes, death of any other cause, e.g. a car accident, can be considered as a competing risk. Both loss to follow-up and competing risks can lead to an underestimation of the risk, because the subjects leaving the study are not any longer able to experience the event of interest and will therefore not be able to add to the numerator in the formula.

Also in the case of a dynamic population, calculating the risk is frequently impossible. Whereas in a 'closed' cohort all measures of disease occurrence can be applied,

it is problematic to measure risk directly in an 'open' cohort, where new people are added during the follow-up period [4]. These situations illustrate that it in many cases it is better to choose an alternative approach to express incidence, i.e. the incidence rate.

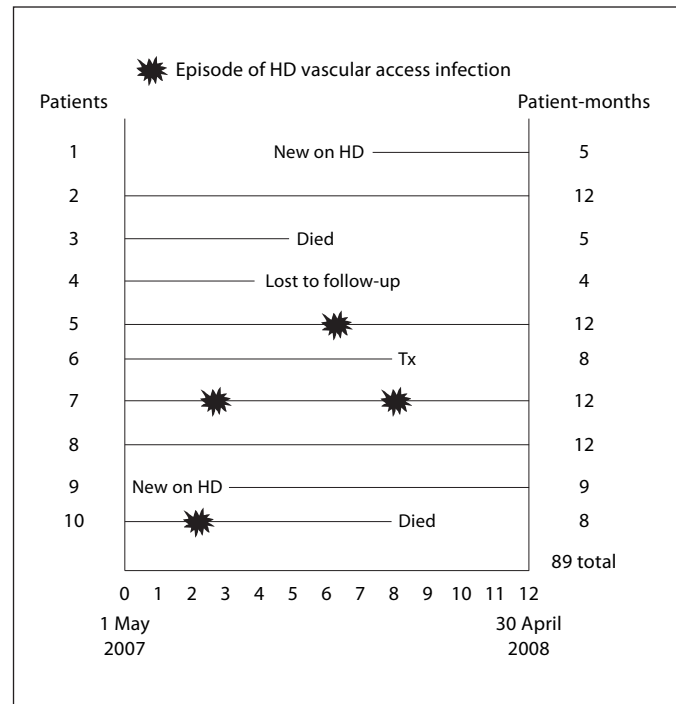
### Incidence Rate

The second measure of disease frequency that expresses incidence is the incidence rate. It can be calculated by dividing the number of subjects developing a disease by the total time at risk for all people to get the disease. The denominator of this formula includes a measure of time instead of just a number of subjects. The incidence rate should therefore be interpreted as an instantaneous concept, like speed:

$$\text{Incidence rate} = \frac{\text{Number of subjects developing the disease}}{\text{Total time at risk for the disease for all subjects followed}}$$

Like for the risk, one assumes for the calculation of the incidence rate that all subjects are free of the disease of interest at the start of the study. However, an important advantage of the incidence rate over the risk is that it is not required for subjects to complete the total follow-up time and only the actual time at risk is taken into account. Figure 1 shows an example of the calculation of the incidence rate. Suppose we study the incidence rate of vascular access infection in haemodialysis (HD) patients in a year and we would have diagnosed a number of episodes of vascular access infection in 10 HD patients. We would then need to calculate the total time at risk; in this case the total time on HD. Figure 1 shows that the 10 patients together were at risk for 89 patient-months. In this period, there were 4 of such episodes. The incidence rate of HD vascular access infection would therefore be  $4/89 = 0.045$  per patient-month, or  $4/7.42 = 0.54$  per patient-year. Assuming that each HD station in a dialysis department is occupied during the full year (by predecessors and successors of the patients shown in fig. 1), one could also calculate the incidence rate of vascular access infection on the level of HD station instead of on patient level. In this case, one would simply need to divide the total number of vascular access infections by the time when the HD stations were occupied (1 year per station).

In larger studies the incidence rate is presented similarly, as is shown in the following example: in 1999, Chow et al. [5] published a study on the rising incidence of renal cell cancer by age, gender, and race in the United States. The authors collected data of patients diagnosed as having kidney cancer between 1975 and 1995 in 9 geographical areas covered by tumour registries. They reported in-



**Fig. 1.** Example of calculating the incidence rate: time at risk for vascular access infection in 10 HD patients. Tx = Renal transplantation.

cidence rates for renal cell carcinoma in white men, white women, black men and black women of 9.6, 4.4, 11.1 and 4.9/100,000 person-years, respectively.

Under conditions in which rates do not change with time (a steady state), the incidence rate can be interpreted as the reciprocal of the average time until an event occurs, also called the waiting time. For example, in the calculation of the incidence rate of vascular access infections in HD patients, the average waiting time for such an episode to occur would be  $1/0.54 = 1.85$  years.

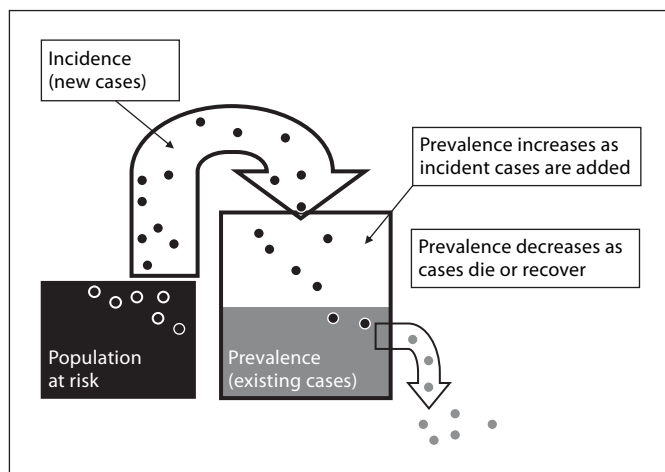
When calculated over a short period of time, the risk and the incidence rate will be rather similar, because the influence of loss to follow-up and competing risks which may flaw risk will only be small.

### Incidence versus Prevalence

Factors that influence the prevalence are the number of incident cases, the deaths, and the recoveries, as is depicted in figure 2. Given a steady state, the prevalence approximately equals the product of the incidence rate and the mean duration of disease. This can be illustrated by

**Table 1.** Summary of characteristics of measures of disease frequency

	Prevalence	Incidence	
Represents	Existing cases at a time point	New cases over a period	
Use	<ul style="list-style-type: none"> <li>- Reflects disease burden</li> <li>- Can be used for planning of health care facilities</li> </ul>	<ul style="list-style-type: none"> <li>- Assessment of disease aetiology</li> <li>- Identification of risk factors</li> </ul>	
		Risk	Incidence rate
Synonyms	Prevalence proportion	Cumulative incidence	Incidence density
Range	0–1 (0–100%)	Incidence proportion	Hazard
Interpretation	Proportion	0–1 (0–100%)	0 to infinity
		Probability	Reciprocal of waiting time



**Fig. 2.** Relationship between incidence and prevalence.

the examples of tetanus and ESRD. Tetanus is an acute, rare, and often fatal disease caused by the bacteria *Clostridium tetani*, leading to rapid death resulting in short disease duration. As a result, its prevalence in the general population will be extremely low at any point in time. ESRD, on the other hand, has a relatively low incidence rate, but in comparison to tetanus its survival is much higher, at least in developed countries. Its average disease duration is much longer; therefore, its prevalence is much higher compared to that of tetanus. Accordingly, an increase in prevalence could be the consequence of a higher incidence of ESRD, of an improved survival, or of a combination of both.

In table 1 the properties of all described measures of disease frequency are summarized.

### Measures of Effect

Once computed, measures of disease frequency may also be used to study the association between exposures and outcomes. The effect of a certain exposure can be studied by comparing the disease frequency in exposed subjects to the disease frequency in those who were not exposed. The comparison of these frequencies can be summarized in a single parameter that estimates the association between the exposure and the disease. This can be accomplished by calculating either the ratio of the measures of disease frequency for two populations which indicates how much more likely one population is to develop a disease than another (relative risk), or the difference between the frequencies which indicates how much greater the frequency of a disease is in one population compared with the other (risk difference). The next paper in this series will address these different measures of effect.

### References

- 1 Dawber TR, Moore FE, Mann GV: Coronary heart disease in the Framingham Study. *Am J Public Health Nations Health* 1957;47:4–24.
- 2 Zelmer JL: The economic burden of end-stage renal disease in Canada. *Kidney Int* 2007;72:1122–1129.
- 3 Ojo AO, Govaerts TC, Schmouder RL, Leichtman AB, Leavey SF, Wolfe RA, Held PJ, Port FK, Agodoa LY: Renal transplantation in end-stage sickle cell nephropathy. *Transplantation* 1999;67:291–295.
- 4 Rothman KJ: *Epidemiology: An Introduction*. New York, Oxford University Press, 2002.
- 5 Chow WH, Devesa SS, Warren JL, Fraumeni JF Jr: Rising incidence of renal cell cancer in the United States. *JAMA* 1999;281:1628–1631.