

## 2 Understanding Basic Epidemiology

### 2.1 Epidemiology

Epidemiology literally means the study of 'what is found amongst the population' (from the Greek: επι δῆμος (epi demos): amongst the population). Well, a lot things are found among the population, like greed, singing talent and typical eating habits, however in this case we mean: diseases and their patterns. These patterns can be described in time as well as in space (months, years, the whole country, provinces) and also in age groups and in gender. Diseases can then be further distinguished after professions, social classes, religions or genetic background, or after urban and rural areas.

Epidemiology is a term minted in the time before germs were distinguished. The science involves all diseases and disorders such as cancer, heart and vascular diseases as well as accidents. It is also very common to speak of epidemiology in terms of suicide or 'child abuse'. A so called 'clinical epidemiology' has developed now alongside general epidemiology. This science aims at the improvement and innovation of clinical treatment and medicine. It studies groups of patients who receive existing or new medicine or treatment, along with placebos. These 'trials' are carried out in a clinical setting.

An **epidemic** is the outbreak of a disease at a larger scale than normally observed. How great the magnitude of the scale has to be before calling it an epidemic depends on the described disease. Some disorders are already called an epidemic at an increase of several percentages in a small amount of time (like rabies). Others, which show a duplication in a short time, are not found alarming (flu).

The word *endemic* is used to describe a disease occurring within a specific area; chicken pox is therefore an endemic disease in The Netherlands. With a not too big increase of some diseases it is called an endemic rise (for example meningococcal meningitis in The Netherlands ± 1990). These terminologies are used for people, animals and plants.

The definition of an *epidemic* by the American Public Health Service is: "The occurrence in a community or region of a group of illnesses (or an outbreak) of similar nature, clearly in excess of normal expectancy and derived from a common or propagated source". Be aware of the lack of the words 'infectious' or contagious! 'Source' can also, for example, be fluorine in water or the presence of additives in food.

Another few terminologies: a *pandemic* is an epidemic that spreads over (a big part of) the world. An example is the Spanish Fever in 1918-1920 and AIDS in our time. A *zoonosis* is a disease that people obtain from animals (for example from cattle or pets). This terminology

nearly always involves contagious diseases. Examples are psittacosis (parrot fever) and Malta fever (from the goat).

Epidemiology is a quantifying science. Doctors focus in general on the individual, while the epidemiologist studies health, illness and mortality in groups. That is why he has to calculate a lot, and especially has to provide a firm basis for his reports: a clear definition of the characteristics of the group and a precise determination of the numbers per group. This is what we call the epidemiological denominator. It may seem like easy given data; in fact it is the most difficult aspect of the profession. The epidemiologist who wants to determine this year's mortality rate for the intravenous drug abusers in Amsterdam cannot easily find an available denominator. The epidemiologist, who wants to determine the annual mortality of road users, will not have sufficient data from death figures alone; he has to distinguish between motorists, cyclists, pedestrians, etc. Also, he has a problem to decide whether to calculate with traffic hours or kilometres.

The *epidemiological numerator* is usually a fraction of the denominator, but sometimes it can exceed the denominator. This is why diarrhoea in children in Bangladesh can be 200,000 per 100,000 per year; when each child on average has such an episode twice in a year. This brings us to the following question: how big will the denominator be? If we say that 121 of the 573 children in a village last year had chicken pox, it is difficult to compare it with another village concerning 84 of the 469. In this case a percentage can be used, but very often the epidemiological fraction is converted into a fraction with a manageable denominator.

Usually it is 100 (percentage) or 1,000 (per mille) or 100,000 (especially popular). Also 10,000 and 1 million occur, but the first mentioned three are preferable. For the previous example it comes down to 211/1,000 compared to 179/1,000. AIDS data, for example, are easily compared between European countries because they are always expressed per 100,000 inhabitants.

With this background knowledge we can now look at the terms morbidity, mortality and case fatality rate.

*Morbidity* (morbus (L) = disease) is the number of cases of a certain disease in a defined group (for example per 100,000 inhabitants of a country).

*Mortality* (mortus (L) = death) is the number of deaths of a certain disease in a defined group (for example per 100,000 inhabitants of a country). It should be clear that the morbidity cannot be smaller than the mortality.

The *Case fatality rate (CFR)* is the number of deaths among patients with a certain disease (for example per 100 –percentage - or per 1,000 patients – per mille). The Dutch term is: *letaliteit* (the English term *lethality* is hardly ever used).

The CFR is of course an indicator for the relation between morbidity and mortality. We can express this in a fraction or ratio:

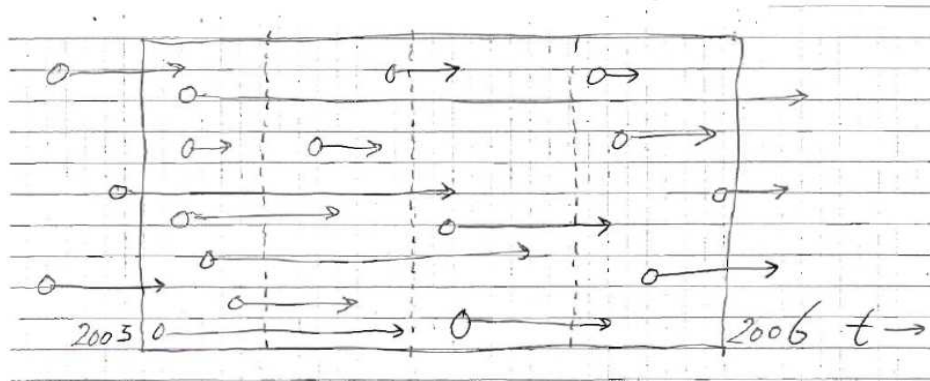
$$\text{Mortality/morbidity} * 100\% = \text{CFR}$$

This expression implies that we not only can calculate CFR from the mortality and morbidity figures. If the CFR of a disease is more or less constant, this parameter can serve to estimate the morbidity provided that the mortality figures are known. In fact this estimation is often used because death due to disease is just better registered than the number of cases of diseases itself.

The description of how many per how many is too inexact for doctors and especially for engineers. The description becomes more accurate for the morbidity, for example, if the dimension of time is taken into account. This gives us the terminologies incidence and prevalence.

*Incidence* is the number of new cases of a certain illness in a given time period (and among a well-defined group, for example 100,000 inhabitants of a country). A year is usually chosen for this period: year incidence. We also have week incidence, month incidence or five-year incidence. This means that cases occurring just at the end of this period are counted in the incidence, but cases that started before the period are not counted, even if the patients are ill the whole period, or even longer.

*Prevalence* is the total number of cases of a certain disease during a given time period (week, month, year, five years) (for example per 100,000). It will be clear that, for example, the year prevalence always is higher than the year incidence. The point prevalence is the most common one. This is the number of patients with a certain disease at a given moment, for example 1 January or 1 July at 00,00 hour. The point prevalence is not always higher than the year incidence. This will be explained now, using the diagram, below.



The point prevalence is much lower than the year prevalence when a disease lasts considerably shorter than one year. An example is the flu. If the disorder lasts about a week, the average point prevalence will be 1/52 of the year incidence. Because season variability exists, it can be 1/20 of the year incidence at a moment in January and 1/200 at a moment in August.

If the illness is chronic and takes much longer than one year, the point prevalence becomes much higher than the year incidence. Examples are illnesses that can last a number of years, like diabetes and multiple sclerosis (MS). If we write the point prevalence as  $P$ , the year incidence as  $I$ , and the duration of the disease as  $D$ , the following formula can be used:

$$P = I \cdot D$$

In this case the duration has to be reproduced as a fraction of a year (for example 0.02 in the case of flu). Also, calculations can be made with two known figures and one unknown figure. For example the incidence will often not be possible to measure, and has to be determined with the point prevalence, linked to a known average disease duration. AIDS is an example of this. By knowing the present number of patients, an estimation can be made of the number of infections that occurred 10 years ago; the (now known) average duration of the incubation time.

Epidemiology is in fact a form of applied population biology. Epidemiological descriptions and analyses can also be made for populations other than humans. The term population means a circumscribed group of individuals, defined by particular features and characteristics. For example, in an investigation of the occurrence of breast cancer in otherwise healthy women aged 50-59 years in the province of North Holland, women with diabetes **and** breast cancer are not enrolled. The same applies for 47-year-old women from the province of Utrecht. The principle is easy, but in practice inclusion and exclusion of individuals provides great problems: for example for clinical epidemiological research.

## **2.2 Medical statistics**

With the background knowledge of the previous parameters, we now can take a look at the medical statistics.

### **Age class**

A group can be divided in a number of consecutive ages (in months or years). Most often five-year classes are used. Watch the notation: 0-4 (0, 1, 2, 3, and 4 years, i.e. until the fifth birthday), 5-9, 10-14, etc. The first two years of someone's life are often divided into months (for instance in the description of infectious diseases like meningitis in young children). Watch out: the first year of one's life is 0-11 months, the second: 12-23 months, the first half year: 0-5 months, the second: 6-11 months.

### **Education classes**

An attempt to describe the material state and the insights of a person.

Well-being and insight (in diseases) are of great importance with respect to contracting diseases or preventing them, but also for diagnosis and treatment. The following groups are usually distinguished: (no) primary school, secondary school (divided into lower and higher levels), and academic education.

### **Genetic group**

In The Netherlands, the genetic subdivision has been neglected until recently. In the U.S. it is often applied and functional. Distinguished, for instance, is Hispanic-White, non-Hispanic White, Black, Native American, and Inuit. Some disorders occur in very different magnitude in several genetic groups, even if a correction is made for well-being and standard of education. Nowadays, the term 'ethnics' is used. It describes a mix of genetic, social and cultural factors that, for instance, can show differences in health and deaths (for example the higher numbers of Moroccan and Turkish children drowning in The Netherlands).

### **Standardization**

The specific age parameters (morbidity, mortality) are reflected onto the corresponding age classes of a standard population, for example, the whole Dutch population. This prevents wrong interpretations. For example, the mortality for pneumonia can seem alarmingly high in a certain area. This can be a misrepresentation if a much higher than normal percentage of people are older than 65 years in that area.

This so called *direct* method can also be reversed: indirect standardization. Then, for example, age specific morbidity figures are projected onto a research population, or index population.

For example in an investigation of tooth cavities, it can appear that a certain population sample has fewer cavities, even after standardization, in comparison to the general population (further investigation can show the higher than average level of fluorine in their water supply).

### **Risk**

The probability of a particular unwanted event occurring within a certain time frame. This probability is not specified towards individual or group characteristics. The chance of a ninety years old (nonetarian) passing away within 10 years is nearly 100 percent. The probability of giving birth to twins is approximately 1.25 percent. Often risk is indicated as a fraction (values 0.0-1.0).

### **Relative risk (RR)**

We distinguish, for instance, relative risk of disease and death. RR reflects the degree to which the probability of a certain event for individuals or groups is greater (or less) than is average for their society. The probability of being involved in a traffic accident is much higher for motorcyclists than for motorists and the chance of getting heart disease is much higher for people with a serum cholesterol of twice the average.

### **Population attributive risk (PAR)**

To allocate the chance of a certain disease or cause of death to a well described *population*. To determine PAR, standardization has to be utilized (such as for ages). The standardization is valued for a certain period and certain criteria are taken into account (for example for traffic: kilometres or hours). It is clear that the safety of a flying population appears better when measured in kilometres rather than in hours when compared with a driving population.

Very often PARs of 1.3 or 1.5 and such are found. However sometimes they can be exceptionally high, e.g. the death risk in a group of climbers of the highest mountain tops, in comparison to the general population between the ages of 20-39 years old. Your lecturer once calculated that the PAR for serious infections by *Streptococcus suis* (a pig's bacterium) was nearly 1500 for slaughterers, butchers and workers in the meat processing industry. This means that IN the total population there are almost no other patients than members of this large professional group.

Although certain people belong to a population with a highly increased probability of a particular disorder or event, this probability can still vary dramatically from person to person within the group. An individual with high serum cholesterol that does not smoke, nor has a high blood pressure or stress, is much less likely to get heart disease than others with high cholesterol levels.

One comes across the RR and the PAR in many epidemiological studies.

### **Disease causes**

The classification of assumed causes of illness (aetiology). Very complex, because often several factors can contribute to the outcome of a disease.

For lung cancer it can be assumed at the moment that 80-90 percent of the cases are due to the smoking of cigarettes.

### **Death causes**

The classification of assumed causes of death. This can be more complex than disease aetiology. Two main categories cause nearly three quarters of the mortality in The Netherlands: heart and vascular disease and cancer.

First something about mortality. Mortality is a “severe” given fact in the medical statistics because it nearly always attracts attention and leads to registration and because it only takes place once per individual; double counts are not easily occurring.

### **Gross mortality figure**

The total number of deaths in a certain period, e.g. the year of 1995 measured out of the total population, or per 1,000. The denominator is taken as the population on 30th June of that year. The gross mortality figure of 1995 was approximately 10 (per 1,000).

### **Specific mortality**

As the previous but now divided into, for example, gender, age groups, training groups, etc. Often then also expressed per 1,000 of that group.

### **Gross mortality after cause**

These mortality figures give the causes of death, which sometimes can be divided further. In 1995 approximately 30 percent of all mortality was due to cancer and 45 percent due to heart and vascular diseases. One quarter of the people that died of cancer died of lung cancer (8/10-9/10 of which was caused by smoking).

### **Specific mortalities after cause**

As the previous, divided into age or training groups, or other parts of the population.

### **Infant mortality rate**

Number of infants dying at the age of 0-11 months. Expressed per 1,000 born alive, usually during the period of one year. In The Netherlands for the year incidence the figure is low, far below 10 (meaning <one percent). In some developing countries it is 100 (10 percent) or even

higher. In The Netherlands the infant mortality rate was also still 10 percent in 1913 (seven in Friesland, thirteen in Brabant).

### **Stillbirth**

Number of children that are born dead after a full-term pregnancy (at least 6 months). Expressed per 1,000 pregnancies. In The Netherlands this value is approximately 10.

### **Neonatal mortality (new born)**

Death during the first four weeks of life (0-27 days). Again per 1,000 and usually during one year.

### **Perinatal mortality**

Child mortality around the birth (the literal meaning of perinatal). This includes the still born and the born alive who die within one week after the birth (0-6 days).

### **Maternal mortality**

Mortality of women by pregnancy complications, abortions, and complications during delivery or period of lying-in. However, when a pregnant woman or a woman who has just given birth dies of another cause, this is not taken into consideration when calculating the maternal mortality. Because the number of pregnancies is difficult to register correctly, it is defined as mortality per 1,000 births (alive or dead).

### **Mortality 0-4 years old**

Mortality amongst infants, toddlers and pre-schoolers. In The Netherlands relatively unimportant (< 1 percent). In some developing countries still between 10 and 20 percent (see chapter demography).

Now a few parameters concerning fertility; the extent to which the population reproduces. These facts are of vital importance for long term planning of employment, schools, training, hospitals, etc.

### **Natality (birth figure)**

Number of new born per year per 1,000 inhabitants. For the denominator, the situation on the 30<sup>th</sup> of June (mid-year population) is taken. In The Netherlands this is around 12-13. In some developing countries the figure can be more than 30, but nearly everywhere in the world it is decreasing.

### **Birth surplus**

The number of living born in a year minus the number of deceased. In The Netherlands it is currently  $\pm 65,000$  (210,000 minus 145,000).



### **Fertility**

The number of born alive calculated over the number of women in the fertile age; more or less in the whole period of 15-44 years old. Usually expressed per 1,000 per year and again the denominator is the situation on 30<sup>th</sup> of June.

Even more important for the size of the human population is *life expectation*. This is expressed in years from birth, but can also be specified per age. It will not surprise anyone that when a girl in a certain country momentarily can look towards 80 years, this number is higher for a now already 90 years old woman. The older you are, the more years in total you can expect. (but the number of remaining years decreases).

The term cohort is connected with this. Historically a cohort was a group of Roman soldiers that belonged together. Whenever a soldier died, he was not replaced in the group. Years later this resulted in only one man walking with the regimental colours. In epidemiology, a cohort of people is e.g. a group that was born in the same year. From that year on individuals disappear by death, until eventually no one is left. In The Netherlands the complete cohort of 1891 has already passed away. And almost everyone of the cohort of 1991 is still alive. The last individuals of the cohort of 2004 will only die in the twenty-second century, around 2115.

### **Life expectation**

The number of years that a cohort can look forward to, based on the mortality figures of that year or from a group of years. On the basis of the current mortality figures, in The Netherlands in 1996, a baby girl could look forward to more than 80 years and a baby boy to more than 75 years. A surprising phenomenon occurs here:

### **Delayed mortality**

The extent to which the expected mortality of a cohort can move to an older age. For the people who were born in the beginning of the 20<sup>th</sup> century in The Netherlands, the expected age at dying was approximately 55 years. During their lifetime it has shifted to more than 75 years. This means a profit of more than 20 years during one's life! This trend was already set, because half way through the 19<sup>th</sup> century the life expectation was only just over 35 years. In China life expectancy has increased in 25 years (1970-1994) by more than 13 years; more than a half year per year!

### **Excess mortality**

Excess of mortality cases with regard to the expected number in a certain period. Well known is the example of the London smog in the beginning of the early fifties, but also in The Netherlands today when there is a big flu epidemic. Another well known example is the hunger winter in the West of The Netherlands in the last year of World War II (1944-1945).

### **Disability Adjusted Life Year (DALY)**

DALY's and DALE's (Disability Adjusted Life Expectancy) were introduced by, for one, the World Bank to make possible comparisons within a country and between countries. The terms concern lost and won life years, corrected by life time spent in a not completely good health. If a person wins, for example, 20 life years, which when corrected has a benefit of only 80 percent at full health, then this can be converted to 16 won years. Turned around, if 10 years are lost, which when corrected counts for 50 percent at full health, the actual loss is only 5 healthy years. These estimations of course have a subjective character. DALY's and DALE's are absolutely not meant to finish with people's wishes, expectations and willingness to expenses. They give epidemiologists the possibility to weigh things up against each other. In table we see these terminologies.

### **Median mortality age**

The age by which 50 percent of the cohort has died. An important parameter, usually when people speak off the *average* mortality age they mean median mortality age. It will be clear that when mortality is high amongst 0-4 year olds in a country, the median life duration will be low, at around 30 years. For those that have survived the first five years, it can however be much better, possibly 50 or 60 years more (so aged 55 or 65).

### **Proportional death rate**

The death of persons above 50 years divided by the total death tally (usually per year). In The Netherlands this is around 90 percent, but in some developing countries it is below 40 or even 30 percent.

### **FIVE years survival**

The percentage of patients of a cohort that survives after, for example, the diagnosis, an operation, or radiotherapy for at least five years. After the diagnosis of lung carcinoma the life expectation is about another half a year; the five years survival is pretty much nil. The diagnosis of AIDS meant till 1996 a life expectation of 2-3 years in The Netherlands, the five years survival was about 10 percent. At the moment – thanks to combination therapy – the majority survives the diagnosis by 5 years and more.

The term can also be used for things other than illness if necessary. The five years survival of a child born alive was less than three quarters in the half of the last century in our country, now it is more than 95 percent.

### **Man to woman ratio**

The relation between the numbers of men and women concerning a certain phenomenon. For example, the incidence of an infectious disease, death by accidents, or amongst infants. Expressed as a fraction (for example 53/47 or 113/100) or in decimals (1.13). However, the

value requires correction depending on the age group concerned. For instance at birth we have approximately 105 boys for every 100 girls. Probably there is still a larger excess of male embryo's during the early stages of the pregnancy, but these are selectively going astray. This phenomenon of higher male mortality continues however after birth in *all* age groups. This means:

- A. A point of equilibrium appears (now between 40 and 50 years)
- B. Amongst elderly people there are more women (a quarter more above 65 years)
- C. The life expectancy of men is five or six years lower. Men switching to female sex by operations and hormones will be rewarded to some degree by a gain in life years.