

# A brief review of epidemiology

Gavin Shaddick

Departments of Mathematical Sciences, University of Bath, UK

A *rate* is defined as the number of events, for example deaths or cases of disease, per unit of population, in a particular time span. To calculate a rate we require:

- ▶ a defined period of time
- ▶ a defined population, with an accurate estimate of the size of the population during the defined period
- ▶ the number of events occurring over the period.

$$\text{rate} = \frac{\text{No. events}}{\text{total person-time at risk}}$$

The *incidence rate* refers to the number of new cases of a particular disease that develop during a specified time interval. For a fixed time period  $\Delta t$ , a population at risk of size  $N$  and the number of new cases of disease  $A$  the incidence rate is

$$\text{incidence} = \frac{A}{N \times \Delta t}$$

An incidence rate lies between 0 and  $\infty$ , and a yearly incidence rate measures the number of cases per person-year.

Measuring the population at risk may be difficult because of changes in the population over the time period. Since it may not be possible to measure disease-free periods precisely, we often calculate the incidence using the average size of the population. This is reasonably accurate if the population size is stable and the incidence rate is low.

The *prevalence* refers to the number of cases of disease that exist at a specified point in time. It is the proportion of the population who have a disease at a given time. Prevalences may also be calculated over a time period; for example the number of events within a time period.

$$\text{prevalence} = \frac{\text{Number of cases}}{\text{Number at risk}}$$

Diseases with high incidence rates may have low prevalences if they are rapidly fatal.

The *crude mortality rate* is usually calculated as deaths per 1000 population per year. Let  $D$  be the number of deaths in a given time period of length  $\Delta t$ , and  $\bar{N}$  be the average size of the population at risk during that period (often approximated by the number in the population at the mid-point of the time period). Then the crude mortality rate is given by

$$r = \frac{d}{\bar{N} \times \Delta t} \times 1000$$

**Table:** Mortality Data by Age Group for England 2003. *Source: ONS Mortality Statistics Series DH1 no. 36*

Age Group	Population (1000s)	% in age group	Deaths	Age-specific mortality rate
0-4	2848.2	5.7	3682	1.3
5-14	6299.8	12.6	725	0.1
15-24	6304.0	12.6	2634	0.4
25-44	14485.7	29.1	14001	1.0
45-64	11971.3	24.0	62755	5.2
65-74	4158.6	8.3	87714	21.1
75+	3788.3	7.6	331898	87.6
All	49855.9	100	503409	10.1

When comparing populations, we can eliminate the effects of, for example, different age structures by looking at age-specific rates. However, this can be cumbersome, and it is often easier to compare a single summary figure.

For indirect standardisation, we take the age-specific rates from the reference population and convert them into the mortality rate we would observe if those reference rates were true for the age-structure of the population of interest.

We calculate the age-specific mortality rates  $r'_k$  for the reference population. Suppose the population of interest has population counts  $N_k$ ;  $k = 1, \dots, K$  in each age-group  $k$ . The expected rate of deaths in the population of interest is

$$\text{expected rate} = \frac{\sum_{k=1}^K N_k r'_k}{\sum_{k=1}^K N_k}$$

and the expected number of deaths in the population of interest is

$$E = \sum_{k=1}^K N_k r'_k$$

We can compare the expected number of deaths, using the indirect standardisation method, with the observed number using the *standardised mortality ratio* (SMR). Let  $O$  be the observed number of deaths in the population of interest, and  $E$  be the expected number of deaths when indirectly standardised with respect to some reference population.

$$\text{SMR} = \frac{O}{E} \times 100$$

The SMR is a ratio, not a rate or a percentage. An SMR of 100 means that the population of interest has the same number of deaths as we would expect from the reference population. If it is greater than 100, we have more deaths than expected; if it is less than 100 we have less.



Let  $D$  and  $\bar{D}$  denote 'disease' and 'not disease', and  $E$  and  $\bar{E}$  denote 'exposed' and 'unexposed' respectively. Then the data for a comparative cohort study, or a case-control study can be written as in Table 2. In a population study, the total sample size  $n_{++}$  is fixed; in a cohort study, the margins  $n_{+0}$  and  $n_{+1}$  are fixed; and in a case-control study, the margins  $n_{0+}$  and  $n_{1+}$  are fixed.

**Table:** Notation for cohort or case-control study.

		<b>Exposure to Risk Factor</b>		
		Unexposed ( $\bar{E}$ )	Exposed ( $E$ )	
<b>Disease</b>	Absent ( $\bar{D}$ )	$n_{00}$	$n_{01}$	$n_{0+}$
	Present ( $D$ )	$n_{10}$	$n_{11}$	$n_{1+}$
		$n_{+0}$	$n_{+1}$	$n_{++}$

The risk of an event is the probability that an event will occur within a given period of time. The risk of an individual having a disease ( $P(D)$ ) is estimated by the frequency with which that condition has occurred in a similar population in the past, that is its incidence:  $n_{1+}/n_{++}$ .

To compare risks for people with and without a particular risk factor, we look at the ratio. Suppose the risk for the exposed group is  $\pi_1 = P(D|E)$  and the risk for the unexposed group is  $\pi_0 = P(D|\bar{E})$ . Then the *relative risk* is

$$RR = \frac{\pi_1}{\pi_0}$$

For a cohort study, we have

$$RR = \frac{n_{11}/n_{+1}}{n_{10}/n_{+0}} = \frac{n_{11}n_{+0}}{n_{+1}n_{10}}$$

An alternative to the risk is the *odds* of an event. If the probability of an event is  $p$  then the odds is

$$\frac{p}{(1 - p)}$$

This can be useful since it is not constrained to lie between 0 and 1. We often use the *log odds*:

$$\log \left( \frac{p}{(1 - p)} \right)$$

Another way to compare people with and without a particular risk factor is to use the *odds ratio*. The odds ratio for disease given exposure is

$$\text{OR} = \frac{\text{odds of disease given exposed}}{\text{odds of disease given unexposed}} = \frac{P(D|E) (1 - P(D|\bar{E}))}{P(D|\bar{E}) (1 - P(D|E))}$$