Chapter 2
Estimates of Associations

Incidence rates and prevalence proportions are used to describe the frequency of diseases and health events in populations. They are also used to estimate an *association* between putative determinants, exposures, and a disease. Epidemiologists often use the term *exposures* to describe a broad range of events, such as stress, exposures to air pollution or occupational factors, habits of life (such as smoking), social conditions (such as income), or static conditions (such as genetic factors). The term, exposure, is thus used to describe all possible determinants of diseases. We are interested in estimating if, and if so, how strongly these exposures are associated with a disease (increase and decrease). We do that by comparing disease frequencies in exposed and unexposed people.

In a simple situation we may observe exposed and unexposed people for a number of months (observation months), and we count newly diagnosed patients in that time. If we assume complete follow-up for 1 year and obtain the data ($N =$ the number of people being followed up, $D =$ disease) of Table 2.1), then one measure of association (under certain strong conditions an estimate of the effect of the exposure for the disease under study) would be the *relative risk*, RR:

$$\mathrm{RR} = \frac{200/1{,}000}{100/1{,}000} = 2.0; \quad \mathrm{RR} = \frac{\mathrm{CI}_{+}}{\mathrm{CI}_{-}}$$

The interpretation is that the estimated risk (CI, cumulative incidence) of getting the disease in the year where we had all in the population under observation (no loss to follow-up) was twice as high among the exposed as it was among the unexposed and there could be many reasons for that.

Another measure of association is the *incidence rate ratio* (IRR):

Table 2.1 Follow-up study with complete follow-up

Exposure	*N*	*D*	Observation years
+	1,000	200	900
–	1,000	100	950

J. Olsen et al., *An Introduction to Epidemiology for Health Professionals*,
Springer Series on Epidemiology and Health 1, DOI 10.1007/978-1-4419-1497-2_2,

$$\text{IRR} = \frac{200/900 \text{ years}}{100/950 \text{ years}} = 2.01; \quad \text{IRR} = \frac{\text{IR}_+}{\text{IR}_-}$$

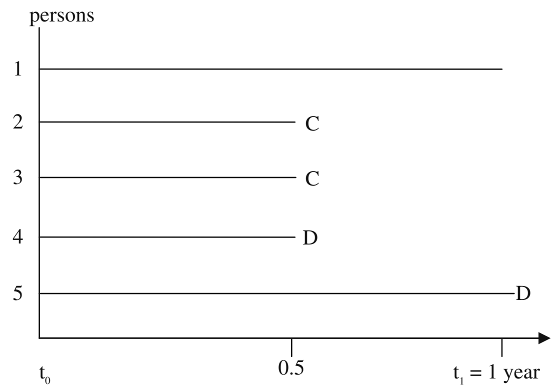
With this measure we state that the incidence rate (IR) of developing the disease per year (new cases per year of observation time among the population at risk) is 2.01 times higher for exposed than for unexposed. Note that this measure does not require complete follow-up of the cohorts.

We may also take an interest in getting an absolute measure of the difference in incidence among exposed compared with unexposed. The *risk difference* or *cumulative incidence difference* will be obtained by subtracting the two cumulative incidences (200/1,000 – 100/1,000) = 0.10. The rate difference will be (200/900 years – 100/950 years) = 0.117 years^{-1}. Relative terms describe how many times the incidence rates for unexposed is to be multiplied to obtain the incidence rate among exposed. The differences provide estimates on an absolute scale. The risk is increased by 10% and the average incidence rate per year is increased by 0.117 years^{-1}.

Notice that these relative and absolute measures of association are purely descriptive. They may, under certain conditions, estimate the effect of exposure, but unless strict (and rare) conditions are fulfilled, the terminology should not promise more than is justified. We are usually interested in effects, but we measure associations. In fact, we are never able to measure effects, only to estimate them.

Usually we have incomplete follow-up even in a fixed cohort because some people leave the study for a number of reasons. They may move out of the area we have under observation (be censored), or they may die from a disease different from the one we study (be censored). Imagine a small segment of our population follow this pattern (D = the disease under study and C = censored observation). If we stop the observation at t_1, we may get the pattern seen in Fig. 2.1.

Fig. 2.1 Observation time



We have two diseased in our population of five people, but only two of the five were under observation for 1 year (1 and 5). An estimated CI of 2/5 = 0.40 may be too low since 2 and 3 could become diseased after they left our study. A CI of

2/3 = 0.66 would be too high – we did observe 2 and 3 for 6 months and they had not been diagnosed with the disease of interest D up to that time. We can, however, use all available information by estimating the incidence rate: 2/(1 + 0.5 + 0.5 + 0.5 + 1) years = 0.571 years^{-1}.

Knowing the incidence rates makes it possible to calculate CI under certain conditions by means of the *exponential formula*

$$\mathrm{CI} = 1 - \mathrm{e}^{-\mathrm{IR} \times \Delta t}$$

In this case we get $1 - \mathrm{e}^{-0.571} = 0.435$ ($\Delta t = 1$), given the incidence rate is constant over the time period (Δt).

The risk of getting the disease over a period of 1 year is 43.5%, but this risk is subject to substantial random variation due to small numbers.

Usually, the incidence rate will not be stable over time, especially if time is age. In that case, we have to stratify the IRs over time intervals, Δ_i, where they are proximally constant, and the formula for using incidence rates to calculate risk becomes

$$\mathrm{CI} = 1 - \mathrm{e}^{-\Sigma_i \mathrm{IR}_i \times \Delta_i}$$

If the disease is rare, like most cancers, the CI is close to $\Sigma_i \mathrm{IR}_i \times \Delta_i$. The risk of getting lung cancer if you live to be 70 is approximately equal to the sum of incidence rates for the age groups 0–9, 10–19, 20–29, 30–39, 40–49, 50–59, and 60–69, multiplied by 10 for these age intervals.

For males (and females) the incidence rates of most cancers are close to 0 up to the age of 30. Let us then say the incidence rates of lung cancer for males per 100,000 observation years are: 0 (0–29), 0.1 (30–39), 0.8 (40–49), 1.2 (50–59), and 3.5 (60–69). The cumulative incidence rates up to age 70 would then be: $0.1 \times 10 + 0.8 \times 10 + 1.2 \times 10 + 3.5 \times 10$ per 100,000 years = 56 per 100,000 observation years, rather close to CI = 1 – e $[(0.1 \times 10 + 0.8 \times 10 + 1.2 \times 10 + 3.5 \times 10)/100{,}000]$:

$$\mathrm{CI} = 0.0005598 \text{ or } 55.98 \text{ per } 100{,}000 \text{ observation years}$$

Incidence rates and incidence rate ratios are what we normally have to measure since we rarely have the opportunity to follow a closed population over time with no censoring, and rates may often be the measure of choice.

http://www.springer.com/978-1-4419-1496-5

An Introduction to Epidemiology for Health
Professionals

Olsen, J.; Christensen, K.; Murray, J.; Ekbom, A.

2010, XII, 163 p., Hardcover
ISBN: 978-1-4419-1496-5