# 6
# Time

## 6.1 When do we start the clock?

In Chapter 5 we discussed the variation of rates with time. In that discussion, by assuming that all subjects entered the study at time zero, we implicitly interpreted time to mean time since entry into the study. However, there are many other ways of measuring time and some of these may be more relevant. For example, in epidemiology, it is usually important to consider the variation of rates with age, for which the origin is the date of birth, or with time since first exposure, for which the origin is the date of first exposure. Similarly, in clinical follow-up studies, time since diagnosis or start of treatment may be an important determinant of the failure rate. In different analyses, therefore, it may be relevant to start the clock at different points. Some possible choices for this starting point are described in Table 6.1.

## 6.2 Age-specific rates

Age is an extremely important variable in epidemiology, because the incidence and mortality rates of most diseases vary with age — often by several orders of magnitude. To ignore this variation runs the risk that comparisons between groups will be seriously distorted, or *confounded*, by differences in age structure.

The assumption that rates do not vary with age can be relaxed by dividing the age scale into bands and estimating a different *age-specific* rate in each band. If the follow-up period is short, so that the age of a

**Table 6.1.** Some time scales

| Starting point | Time scale |
|---|---|
| Birth | Age |
| Any fixed date | Calendar time |
| First exposure | Time exposed |
| Entry into study | Time in study |
| Disease onset | Time since onset |
| Start of treatment | Time on treatment |

**Table 6.2.** Entry and exit dates for the cohort of four subjects

| Subject | Born | Entry | Exit | Age at entry | Outcome |
|---|---|---|---|---|---|
| 1 | 1904 | 1943 | 1952 | 39 | Lost |
| 2 | 1924 | 1948 | 1955 | 24 | Failure |
| 3 | 1914 | 1945 | 1961 | 31 | Study ends |
| 4 | 1920 | 1948 | 1956 | 28 | Unrelated death |

subject does not change appreciably during follow-up, age-specific rates can be estimated by classifying subjects into age *groups* by their age at entry. Each subject appears in only one age group and a separate rate is estimated for each group. For longer studies it will be necessary to take account of changing age during the study, and to treat age properly — as a time scale. This scale is then divided into *bands* and a separate estimate of the rate is made within each age band as described in Chapter 5. In this latter analysis, a subject can pass through several age bands during the course of the study.

To see how the failures and observation time are divided between age bands consider the cohort of four subjects, shown in Table 6.2. Subject 1 is lost to follow-up in 1952, subject 2 fails in 1955, subject 3 is still under observation when the study period ends, and subject 4 dies from an unrelated cause in 1956. The date when a subject joins the cohort is called the entry date and the date when observation stops, for whatever reason, is called the exit date. The time between the entry and exit dates is the observation time for the subject. To simplify the exercises, we give dates only as years and will assume that all events take place on the first day of the year. In practice, times would be worked out as accurately as the data allow.

**Exercise 6.1.** What are the observation times for the members of this cohort?

Figure 6.1 shows the observation of the subjects in calendar time, while Figure 6.2 shows it on a scale where time is measured from each subject's date of birth. To estimate a rate for a particular age band the failures are allocated to the bands in which they occurred, and the observation time is divided according to how long the subjects spend in each of the age bands. For example, the age band 30-34, which is from exact age 30 to just less than exact age 35, contains one failure and 10 person-years of observation time, so the estimated rate is 1/10 per person-year.

In this example the observation times in the different time bands have been obtained from the figure, but in practice the total observation time in an age band is obtained by using the dates when the subject changes age bands. For example, subject 1 is 39 years old on entry so he starts in the age band 35–39. He changes age band in 1944 (when he is 40), and again in 1949 (when he is 45), and he leaves the study in 1952 (when he



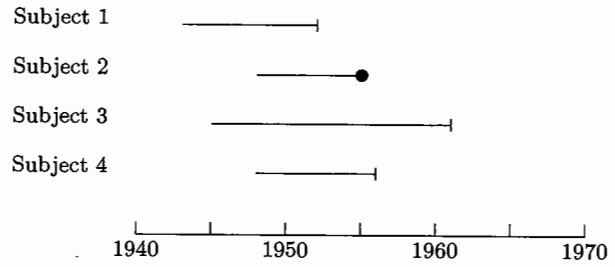


**Fig. 6.1.** Follow-up of four subjects by calendar time.

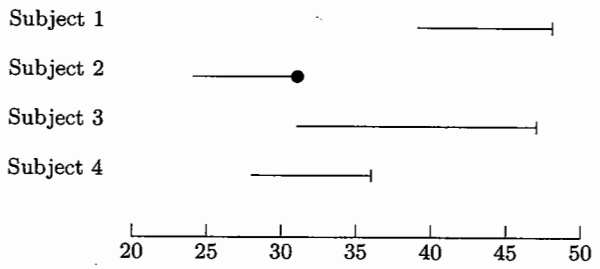


**Fig. 6.2.** Follow-up of four subjects by age.

emigrates). The observation time he spends in the different age bands is shown in Table 6.3.

As a check, the total observation time for subject 1 is from 1943 to 1952 which is 9 years, equal to the sum of the separate times spent in the different age bands.

**Exercise 6.2.** Subject 5 is born in 1931, joins the cohort in 1953, and is lost to follow-up in 1957. Divide the observation time for this subject between the five-year age bands shown in Figure 6.2.

**Table 6.3.** Time in each age band for subject 1

| Age band | Date in | Date out | Time |
|---|---|---|---|
| 35–39 | 1943 | 1944 | 1 |
| 40–44 | 1944 | 1949 | 5 |
| 45–49 | 1949 | 1952 | 3 |

**Table 6.4.** Woman-years and reference rates for a breast cancer study

| Age | Woman-years | E & W rate per 100 000 woman-years |
|---|---|---|
| 40–44 | 975 | 113 |
| 45–49 | 1079 | 162 |
| 50–54 | 2161 | 151 |
| 55–59 | 2793 | 183 |
| 60–64 | 3096 | 179 |

## 6.3 The expected number of failures

One reason for subdividing the total follow-up experience of a cohort into age bands is to determine whether the observed number of failures is more or less than we might have expected. Since mortality and incidence rates usually increase quite sharply with age, the distribution of person years observation between age bands is an extremely important determinant of the number of events we would expect to observe.

Table 6.4 shows the partition of woman-years between age bands for a cohort study of 974 women given a hormone treatment at menopause. During the follow-up period, 15 new cases of breast cancer occurred in the cohort. We might ask whether this is more or less than we would expect from national rates.

The third column of the table shows the age-specific incidence rates of breast cancer for England and Wales at the time the study was carried out. If the rates in the study population are the same as in the rest of England and Wales, the number of cases we would expect in each age band is simply the product of the woman-years observation and the rate. Thus, for the 40–44 age band, the expected number of cases is

$$975 \times \frac{113}{100\,000} = 1.10.$$

**Exercise 6.3.** Carry out these calculations for the remaining age groups and calculate the total expected number of cases of breast cancer.

This exercise shows that 16.77 cases are expected from national rates using the person years in the study. This expected number of cases is quite close to the observed 15, so that there is little suggestion that the rates in this cohort are unusual.

The expected number of cases, as calculated above, is not quite the same as the expected number in the usual statistical sense. The latter cannot depend upon the outcome of the study, but the former does, since the total person-time of observation in the study varies according to how many subjects fail and when. However, for the rare events studied by epidemiologists, this variation is small enough to be ignored.



## 6.4 Lexis diagrams

More than one time scale can be important in the same study. For example, mortality rates from cancer of the cervix depend upon age, as a result of the age-dependence of the incidence rate, and upon calendar time as a result of changes in treatment, population screening, and so on. The situation is further complicated by the strong dependence of the incidence of this disease upon sexual behaviour, which varies from one generation to the next.

The way to separate the effects of two time scales on a rate is to divide each scale into bands, usually of equal width, and to make a separate estimate of the rate for each pairing of bands. To see how this is done in practice it is best to show the subjects relative to the two scales simultaneously, in what is called a *Lexis diagram.*

The four subjects in Table 6.2 are shown relative to both age and calendar year simultaneously in the Lexis diagram in Figure 6.3. Each rectangular region in a Lexis diagram corresponds to a combination of two bands, one from each scale. To estimate rates for these combinations of bands the failures are allocated to the rectangles in which they occur and the observation time for each subject is divided between rectangles according to how long the subjects spends in each.

For example, subject 1 joins the cohort in 1943 aged 39. He changes age bands one year later in 1944 then 5 years later in 1949. He changes calendar periods in 1945 and 1950. Finally, observation stops in 1952. The subdivision of the observation time for this subject between different age and calendar period combinations is shown in Figure 6.4. Note that the times in the different bands add to 9 years, the total observation time for this subject. For each combination of age band and calendar period the rate is estimated by dividing the number of failures by the person-time of observation.

**Exercise 6.4.** Trace the progress of subject 1 through the squares in Figure 6.3 and verify the results given above. Divide the observation time for subject 2 between combinations of five-year bands of age and calendar time in the same way.

The same procedure can be used to separate the effect of age from the effect of time since entry, although there may not be enough data for some combinations of age and time since entry to estimate a rate. Figure 6.5 shows the four subjects in the cohort relative to age and time since entry. Five-year bands have again been chosen for both scales.

**Exercise 6.5.** Divide the observation time for subject 1 between different combinations of five-year bands of age and time since entry.

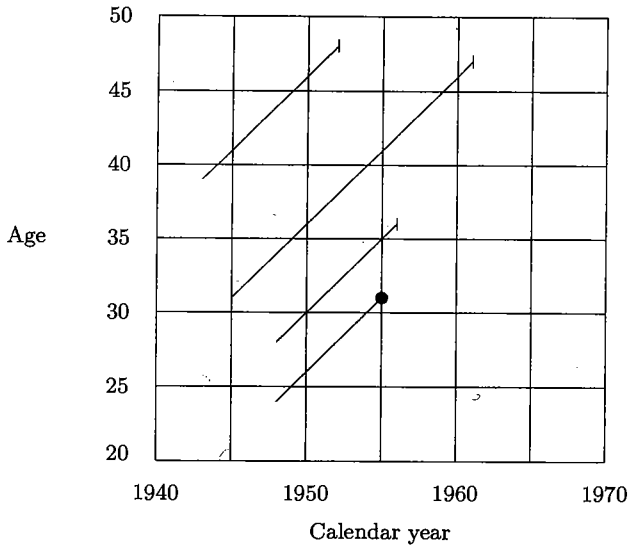


**Fig. 6.3.** Lexis diagram showing age and calendar period.

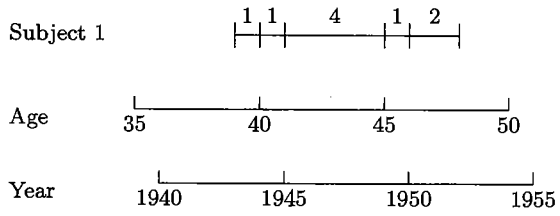


**Fig. 6.4.** Follow-up of subject 1 by age and calendar time.

### 6.5 Reference rates by calendar period

Reference rates, used to calculate the expected numbers of failures, usually come from national rates tabulated by age, sex, and calendar period. In the UK these are calculated using an approximate figure for the person-years. For example, the all-cause mortality rate for the age band 50–54 during 1983 is estimated by $D/Y$ where $D$ is the number of deaths during 1983 for which the subject's age at death was in the range 50–54, and $Y$ is the person-time lived during 1983 by that part of the population whose ages were in the range 50–54 during 1983. Since the exact value of $Y$ is not



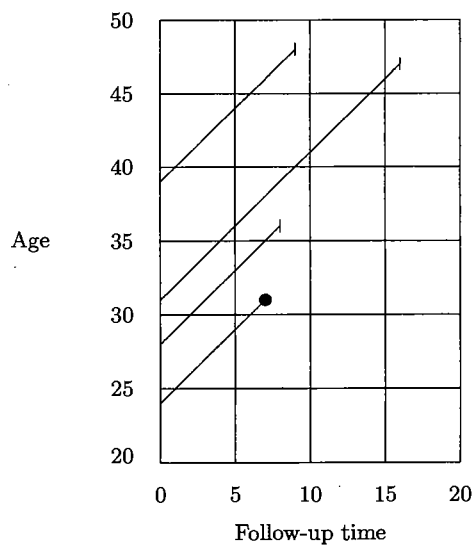


**Fig. 6.5.** Lexis diagram showing age and time since entry.

known an approximate value is obtained from

$$Y \approx \text{Population aged 50–54 in mid-1983} \times 1 \text{ year}.$$

For five-year calendar periods such as 1981–85,

$$Y \approx \text{Population aged 50–54 in mid-1983} \times 5 \text{ years}.$$

The population in the different age bands for any year is obtained from the census; directly for census years and indirectly for inter-census years by updating the last census by births, deaths, and migration.

**Exercise 6.6.** The total number of deaths from cancer of the lung in the SW region of England during the years 1981–88 were males: 14 751, females: 5420. The 1984 population of the region is estimated to be males: 2 154 900, females: 2 306 300. Calculate the mortality rate per $10^6$ person-years for males and females separately.

When follow-up of a cohort takes place over an extended calendar period, the national age-specific rates will usually vary over this period, making it difficult to choose a single set of age-specific rates to use for comparison purposes. The solution is to compute the expected number of events by both age and calendar period, using the appropriate national rates for each calendar time period. To do this the person-years observation in the co-

**Table 6.5.** Mortality following X-irradiation

| Cause of death | Number of deaths | | Ratio |
|---|---|---|---|
| | Observed, $D$ | Expected, $E$ | $D/E$ |
| Cancers: | | | |
| Leukaemia | 31 | 6.47 | 4.79 |
| Colon | 28 | 17.30 | 1.62 |
| Heavily irradiated sites | 259 | 167.50 | 1.55 |
| Lightly irradiated sites | 79 | 65.65 | 1.20 |
| All neoplasms | 397 | 256.92 | 1.55 |
| Other causes | 1362 | 804.68 | 1.69 |
| All causes | 1759 | 1061.61 | 1.66 |

hort study must be partitioned by age and calendar period. The expected number of failures can then be calculated for each combination of age and calendar period, as before, by multiplying the person-years observation by the appropriate national rate. Addition over all combinations of age and calendar period yields an expected number of cases which takes account of variation in national rates with both age and calendar time.

An example of this kind of calculation appears in Table 6.5, which shows some results taken from a study of cancer mortality in a cohort of ankylosing spondilitis patients who had been treated with a single course of X-irradiation of the spine.* The follow-up of each patient started in the year of treatment (1935–1954) and continued until death, migration or 1970 (the date when this analysis was carried out). Follow-up was also terminated by a second course of treatment because the aim was to study the effect of a single course of X-rays and the time before this effect became apparent. The study was carried out in Great Britain and Northern Ireland, and the expected numbers of deaths calculated using the national rates for England and Wales, tabulated by five-year bands for both age and calendar time. It can be seen that mortality from all causes was higher in this cohort than in the reference population. Although accounting for relatively few excess deaths, the *ratio* of observed to expected deaths was particularly high for leukaemia. This ratio is an important index in epidemiology and is called the *standardized mortality ratio* (SMR). We shall discuss it further in Chapter 15.

**Exercise 6.7.** Table 6.6 subdivides the observed and expected deaths from leukaemia according to time since X-ray treatment. How would this table have been calculated?

*From Smith, P.G. and Doll, R.(1982) *British Medical Journal*, **284**, 449–460.



**Table 6.6.** Leukaemia deaths by time since treatment

| | Time since treatment (years) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0–2 | 3–5 | 6–8 | 9–11 | 12–14 | 15–17 | 18–20 | >20 |
| Observed | 6 | 10 | 6 | 3 | 1 | 4 | 1 | 0 |
| Expected | 1.00 | 0.89 | 0.87 | 0.90 | 0.96 | 0.90 | 0.55 | 0.40 |
| Ratio | 6.00 | 11.24 | 6.90 | 3.33 | 1.04 | 4.44 | 1.82 | 0.00 |

## Solutions to the exercises

**6.1** The observation times for the four subjects are 9, 7, 16, and 8 years respectively.

**6.2** Subject 5 is 22 years of age on joining the cohort and 26 when lost to follow-up. She contributes 3 years to the band 20–24, and 1 year to the band 25–29.

**6.3** The expected numbers of cases in the five age bands are 1.10, 1.75, 3.26, 5.11, and 5.54. The sum of these values is 16.76, but working to full accuracy we obtain 16.77 for the total expected number of cases.

**6.4** The Age×Period bands in which subject 2 was observed are as follows:

| Age | Calendar period | Time in band |
|---|---|---|
| 20–24 | 1945–49 | 1 |
| 25–29 | 1945–49 | 1 |
| 25–29 | 1950–54 | 4 |
| 30–34 | 1950–54 | 1 |

**6.5** The Age×Follow-up bands in which subject 1 was observed are as follows:

| Age | Follow-up time | Time in band |
|---|---|---|
| 35–39 | 0–4 | 1 |
| 40–44 | 0–4 | 4 |
| 40–44 | 5–9 | 1 |
| 45–49 | 5–9 | 3 |

**6.6** The estimated rate for males is

$$\frac{14\,751}{2\,154\,900 \times 8} = 856 \text{ per } 10^6 \text{ person-years}$$

and the estimated rate for females is

$$\frac{5\,420}{2\,306\,300 \times 8} = 294 \text{ per } 10^6 \text{ person-years.}$$

**6.7** The follow-up of each subject can be represented by a line on a three-dimensional Lexis diagram with axes: age, period, and time since treatment. Age and period were divided into five-year bands and time since treatment into three-year bands. Observed deaths and person-years can be assigned to cells in the resulting three-dimensional table. Multiplication of person-years by national rates gives the expected number of deaths for each cell. Table 6.6 is formed by adding this table over age and period.