Chapter 8
Cohort studies

A cohort study is an observational study in which a study population (a cohort) is selected and information is obtained to determine which subjects either have a particular characteristic (e.g., blood group A) that is suspected of being related to the development of the disease under investigation, or have been exposed to a possible etiological agent (e.g., cigarette smoking). The entire study population is then followed up in time, and the incidence of the disease in the exposed individuals is compared with the incidence in those not exposed.

Thus cohort studies resemble intervention studies in that people are selected on the basis of their exposure status and then followed up in time, but differ from them in that the allocation to the study groups is not under the direct control of the investigators.

Example 8.1. *A cohort study of 22 707 Chinese men in Taiwan was set up to investigate the association between the hepatitis B surface antigen (HBsAg) and the development of primary hepatocellular carcinoma. The study was conducted among male government employees who were enrolled through routine health care services. All participants completed a health questionnaire and provided a blood sample at the time of their entry into the study. Participants were then followed up for an average of 3.3 years (Beasley* et al.*, 1981).*

In Example 8.1, a group of 22 707 Chinese male government employees (the 'cohort') was assembled and their HBsAg status (the 'exposure') determined at the start of the study. They were then followed up for several years to measure (and compare) the incidence of hepatocellular carcinoma (the 'outcome') in subjects who were HBsAg-positive or HBsAg-negative at the time of entry into the study.

8.1 Definition of the objectives

As in any other study design, it is essential that a clear hypothesis is formulated before the start of a cohort study. This should include a clear definition of the exposure(s) and outcome(s) of interest. Since cohort studies in cancer epidemiology often involve follow-up of a large number of people for a long period of time, they tend to be very expensive and time-consuming. Consequently, such studies are generally carried out after a hypothesis has been explored in other (cheaper and quicker) types of

study (e.g., cross-sectional or case–control studies). For instance, the cohort study in Taiwan (Example 8.1) was set up only after a series of case–control studies of hepatocellular carcinoma had been carried out in the early and mid-1970s (IARC, 1994b).

8.2 Choice of the study population

8.2.1 Source of the study population

The choice of a particular group to serve as the study population for any given cohort study depends on the specific hypothesis under investigation and on practical constraints. The cohort chosen may be a general population group, such as the residents of a community, or a more narrowly defined population that can be readily identified and followed up, such as members of professional or social organizations (e.g., members of health insurance schemes, registered doctors and nurses). Alternatively, the cohort may be selected because of high exposure to a suspected etiological factor, such as a source of ionizing radiation, a particular type of treatment (e.g., chemotherapy, radiotherapy) or an occupational hazard.

A *general population cohort* may be drawn from a geographically well defined area (as in Example 8.2), which is initially surveyed to establish baseline exposure status with respect to a number of factors and then examined periodically to ascertain disease outcomes.

Example 8.2. *A cohort of 10 287 individuals resident in the Ernakulam district of Kerala (India) were followed up for a 10-year period to assess the effect of tobacco chewing and smoking habits on overall mortality. Participants were initially identified through a baseline survey in which a number of villages in the district were randomly selected. All residents in the selected villages aged 15 years and over were interviewed about their tobacco habits in a house-to-house survey and entered into the cohort. Refusals were negligible (Gupta* et al.*, 1984).*

One of the great advantages of this type of cohort study is that it allows a large number of common exposures to be considered in relation to a large number of outcomes. The Framingham Study is a classical example of this. Approximately 5000 residents of the town of Framingham, in Massachusetts (USA), have been followed up since 1948 (Dawber *et al.*, 1951). There were several reasons for selecting this location for the study, mainly determined by logistic and other practical considerations to ensure that it would be feasible to identify and follow participants for many years. At the time the study was set up, Framingham was a relatively stable community including both industrial and rural areas, with a number of occupations and industries represented. The town was small enough to allow residents to come to one central examining facility and there was only one major hospital. Follow-up of this cohort has permitted assessment of the effects of a wide variety of factors (e.g., blood pressure, serum

cholesterol, alcohol intake, physical exercise, smoking) on the risk of numerous diseases, ranging from cardiovascular diseases and cancer to gout, gallbladder disease and eye conditions.

Alternatively, it can be preferable for logistic reasons to draw a general population cohort from a well defined socio-professional group of individuals. For instance, the Taiwan study described in Example 8.1 was conducted among civil servants not because they were thought to have a higher exposure to hepatitis B virus than the rest of the population, but because this group of people was easier to identify and to follow than any other potential study population.

Example 8.3. *A postal questionnaire was sent in 1951 to all men and women whose names were at that time on the British Medical Register and who were believed to be resident in the United Kingdom. In addition to name, address and age, they were asked a few simple questions about their smoking habits. A total of 34 439 male and 6194 female doctors provided sufficiently complete replies. These doctors have been followed up since then (Doll & Hill, 1954; Doll & Peto, 1976; Doll* et al.*, 1980, 1994a,b).*

Similarly, when Richard Doll and Bradford Hill set up a cohort study in England and Wales to assess the health effects of smoking, the choice of the British physicians as the study population (Example 8.3) was determined mainly by logistic considerations. Physicians were registered with the British Medical Association and were therefore easy to identify and follow up. Besides, they were more likely to cooperate and the cause of death to be properly investigated.

If the *exposure is rare*, a study of the general population will have little ability to detect an effect (i.e., the study would have insufficient statistical power (see Chapter 15)), since very few people would have been exposed to the factor of interest. This problem can be overcome by deliberately selecting a *highly exposed group* of people as the study population. For example, exposure to dyestuffs is rare in the general population. However, by choosing a group of workers with high exposure, the full range of effects of the exposure on their health can be studied, including outcomes that are rare in the general population but not in those heavily exposed. The general public health impact of the exposure may be small, but such studies can give insight into common biological mechanisms in disease.

Example 8.4. *The Life Span Study is an on-going cohort study which was set up to investigate the long-term health effects of exposure to high levels of ionizing radiation among survivors of the atomic bomb explosions in Japan. It comprises a sample of 120 128 subjects who were resident in Hiroshima and Nagasaki in 1950, when the follow-up began (Shimizu* et al.*, 1990).*

The follow-up of the survivors of the atomic bomb explosions in Japan (Example 8.4) has not only clarified many of the long-term health effects of acute exposure to high levels of ionizing radiation, but has also contributed to our understanding of the effects of chronic exposure to low-level radiation.

8.2.2 Choice of the comparison group

Once the source of exposed subjects has been determined, the next step is to select an appropriate comparison group of unexposed individuals. This selection is the most critical aspect in the design of a cohort study. The unexposed group should be as similar as possible to the exposed group with respect to the distribution of all factors that may be related to the outcome(s) of interest except the exposure under investigation. In other words, if there were really no association between the exposure and the disease, the disease incidence in the two groups being compared would be essentially the same. Two main types of comparison group may be used in a cohort study: *internal* and *external*.

General population cohorts tend to be heterogeneous with respect to many exposures and, hence, their members can be classified into different exposure categories. In such circumstances, an *internal comparison group* can be utilized. That is, the experience of those members of the cohort who are either unexposed or exposed to low levels can be used as the comparison group. For example, in the cohort study of British physicians (Example 8.3), it was possible to categorize individuals in terms of their smoking habits and then compare the mortality from lung cancer (and other conditions) in smokers with mortality in non-smokers.

Example 8.5. *The Nurses' Health Study was established in 1976, when a cohort of 121 700 US female registered nurses aged 30–55 years completed a questionnaire on medical conditions and lifestyle practices. A total of 1799 newly diagnosed breast cancer cases occurred during the first 10 years of follow-up from mid-1976 to mid-1986. Analyses were then conducted to investigate the relationship between oral contraceptive use and risk of breast cancer. Women who reported in the initial questionnaire in 1976 and in subsequent ones to have never taken oral contraceptives were considered as the 'unexposed' group in this analysis (Romieu* et al.*, 1989).*

In Example 8.5, a portion of the cohort of US registered female nurses was taken as the 'unexposed group' to examine the relationship between oral contraceptive use and risk of developing breast cancer.

In general population cohorts, it is possible to examine the effect of more than one exposure. Thus, the choice of the group of people in the cohort who will be regarded as 'unexposed' depends on the particular exposure under investigation. For instance, the Nurses' Health Study has also allowed examination of the relationship between dietary total fat

intake and the risk of breast cancer. For this purpose, nurses were asked to complete a dietary questionnaire and the distribution of fat intake in the whole cohort was examined and divided into five groups of equal size ('quintiles'); women in the lowest quintile of fat intake were taken as the 'unexposed group' (Willett *et al.*, 1987).

In occupational cohorts, an internal comparison group might consist of workers within the same facility with other types of job not involving exposure to the factor under investigation.

Example 8.6. *A cohort of all 14 282 workers employed at the Sellafield plant of British Nuclear Fuels at any time between the opening of the site in 1947 and 31 December 1975 was identified retrospectively from employment records. Employees who worked in areas of the plant where they were likely to be exposed to external radiation wore film badge dosimeters and these personal dose records were kept by the industry. These workers were considered in the present study as 'radiation workers', while those who never wore film badges were taken as 'non-radiation' workers. It was initially planned to follow up the workers from the time they joined the workforce up to the end of 1975, but the follow-up period was later extended to the end of 1988. The mortality experienced by the 'radiation workers' was then compared with that experienced by the 'non-radiation workers' (Smith & Douglas, 1986; Douglas* et al.*, 1994).*

In Example 8.6, it was possible to identify a group of workers who could be regarded as unexposed to external radiation on the basis of personal dose records.

When the cohort is essentially homogeneous in terms of exposure to the suspected factor, a similar but unexposed cohort, or some other standard of comparison, is required to evaluate the experience of the exposed group. For example, in some occupational cohorts it is not possible to identify a subgroup of the cohort that can be considered as 'unexposed' for comparison. In this instance, an *external comparison group* must be used. A potential comparison group is a cohort of similar workers in another occupation which does not involve exposure to the factor of interest. For instance, many occupational exposures only occur among certain workforces and therefore it can often be assumed that the level of exposure of other workforces is virtually zero. We can therefore choose people in employment from the same geographical area, who are not exposed to the risk factor of interest, as a comparison group. It is important to ensure that the risk of disease in these workforces is not affected by their own occupational exposures.

Alternatively, the general population of the geographical area in which the exposed individuals reside may be taken as the external comparison group. In this case, the disease experience observed in the cohort is compared with the disease experience of the general population at the time the cohort is being followed. Comparison with rates in the general population

avoids the need to follow up a large number of unexposed individuals, but it has several disadvantages. First, it can be done only for outcomes for which such information exists for the general population. Second, it assumes that only a very small proportion of the general population is exposed to the risk factor of interest, otherwise the presence of the exposure in the comparison group will lead to a gross underestimation of its true effect. Third, even if the general population is chosen to be as similar as possible to the exposed cohort in relation to basic demographic and geographic characteristics, it may well differ with respect to other risk factors for the disease, such as diet, smoking, etc. Since this information is not available on individuals in a general population, any observed differences may in fact be due to the effects of confounding that cannot be controlled.

The advantage of using another special group of people as the external unexposed comparison group rather than making comparison with disease rates of the general population is that the group can be selected to be more similar to the exposed cohort than the general population would be. Moreover, information on potential confounding factors can be obtained from all exposed and unexposed individuals in the study and differences controlled for in the analysis.

In many cohort studies, it may be useful to have *multiple comparison groups*, especially when we cannot be sure that any single group will be sufficiently similar to the exposed group in terms of the distribution of potential confounding variables. In such circumstances, the study results may be more convincing if a similar association were observed for a number of different comparison groups. For instance, with some occupational cohorts both an internal comparison group (people employed in the same factory but having a different job) and the experience of the general population (national and local rates) may be used.

In Example 8.7, the all-cause mortality of the cohort of rubber workers was compared with the mortality of another industrial cohort and with local (state) and national rates. Note that both the rubber and the steel workers experienced lower age-specific death rates than either the state or the national populations. This is because people who work tend to be healthier than the general population, which includes those who are too ill or disabled to work (although for steel workers the difference may be due partly to changes in mortality over time). This well known selection bias is called the 'healthy worker effect'.

The healthy worker effect may conceal true increases in the risk of a disease in relation to a particular exposure. It is known to vary with type of disease, being smaller for cancer than for other major diseases, and it tends to disappear with time since recruitment into the workforce (see Section 13.1.1). If rates in the occupational cohort remain lower than those from the general population throughout the follow-up period, this is more likely to be due to sociodemographic and lifestyle differences between the workforce and the general population than to the selection of healthy individuals at the time of recruitment.

Example 8.7. *A cohort of workers in a major tyre-manufacturing plant in Akron, Ohio (USA) was set up to examine their overall and cause-specific mortality. A total of 6678 male rubber workers aged 40 to 84 at 1 January 1964 were identified retrospectively from pension, payroll, death claims and other company files. These workers were followed from 1964 to 1972. The age-specific mortality experienced by this cohort was then compared with that experienced by three comparison groups–an industrial cohort of steel workers, the population of the state where the plant is located (Ohio) and the US national population (Table 8.1) (McMichael* et al.*, 1974).*

	Age-specific mortality rate (per 100 000 pyrs)			
Age-group (years)[b]	**Rubber worker cohort (1964–72)**	**Steel worker cohort (1953–61)**	**Ohio state (1972)**	**USA (1968)**
45–54	852	907	940	980
55–64	2317	2166	2365	2370

[a] Data from McMichael *et al.* (1974).

[b] Only data for these two age-groups were available for all the four populations.

Table 8.1.
Male age-specific mortality rates from all causes in the rubber worker cohort and in three other comparison groups: steel workers, Ohio state population and USA national population.[a]

This health selection effect is not restricted to occupational cohorts. A similar phenomenon has been observed in many other types of cohort study. In the British doctors study described in Example 8.3, those who replied to the initial questionnaire had a much lower mortality in the first years of follow-up than those who did not reply (Doll & Hill, 1954). Less health-conscious people, or those already suffering from health problems, might have felt less motivated to participate.

8.3 Measurement of exposure

Measurement of the exposure(s) of interest is a crucial aspect in the design of a cohort study. Information should be obtained on age at first exposure, dates at which exposure started and stopped, dose and pattern of exposure (intermittent versus constant), and changes over time (see Section 2.3).

Information on the exposure(s) of interest may be obtained from a number of sources, including records collected independently of the study (e.g., medical, employment or union records); information supplied by the study subjects themselves, through personal interviews or questionnaires; data obtained by medical examination or other testing of the participants; biological specimens; or direct measurements of the environment in which cohort members have lived or worked. The advantages and limitations of each of these sources were discussed in Chapter 2.

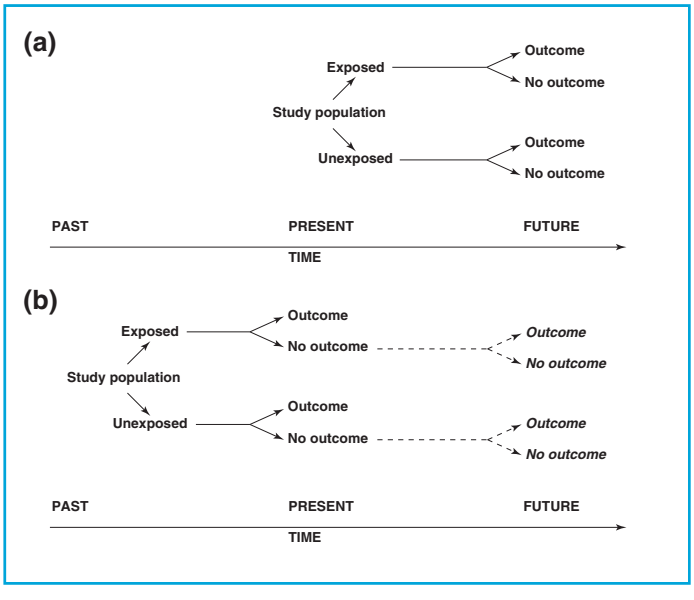


Figure 8.1.
Outline of (*a*) a prospective cohort study and (*b*) a historical cohort study.

There are two main types of cohort study, defined according to the point in time when information on exposure was collected: present or past. In *prospective cohort studies*, data on exposure is collected now, once the study population has been defined. In this instance, it is possible to use the most up-to-date methods of exposure measurement so that bias in exposure classification can be minimized. The main disadvantage of this type of cohort study, however, is that the time from exposure to onset of disease (i.e., the induction period) may be too long (many decades for most cancers). Examples 8.1 to 8.5 are examples of prospective cohort studies which involved the follow-up of large numbers of people for very long periods of time.

The alternative, particularly useful for conditions with long induction periods, is to rely on exposure measurements made many years before the study was set up, which may be available from medical, employment or other personal records. By use of data from existing records, the time we have to wait for the exposure to have any effect on the risk of disease may be considerably reduced or even eliminated. This type of cohort study is called a *historical cohort study*.

Example 8.8. *In the early 1950s, Case and his co-workers set up a cohort study to assess whether men engaged in the manufacture of certain dyestuff intermediates had an excess risk of bladder cancer. They began by constructing a list of all men who had ever been employed in the chemical industry in the United Kingdom for at least six months since 1920. The age and the dates between which exposure to dyestuffs occurred were recorded. A search was made retrospectively for all bladder cancer cases occurring among men who had been employed in the chemical industry, in or after 1921 until 1 February 1952. The number of observed bladder cancer cases among these workers was then compared with the number that would have been expected if these workers had the same mortality experience as the general population of the United Kingdom (Case* et al.*, 1954; Case & Pearson, 1954).*

The study described in Example 8.8 is a classic example of the use of this historical approach. Examples 8.6 and 8.7 are also illustrations of historical cohort studies, since both relied on preexisting employment records to identify the cohort members and to classify them according to their exposure status. Historical cohort studies are particularly useful in occupational epidemiology because, if there is concern that a particular exposure may be a hazard, it is not reasonable to wait decades for clarification in a prospective cohort study. However, if at the time the historical cohort is identified, a large proportion of members are still alive, the follow-up period can be extended into the future (as in Example 8.6) to ensure that all possible long-term health effects are properly assessed.

One of the main limitations of historical cohort studies is that the exposure data available in past records are generally less accurate and detailed

than if they were collected prospectively. Thus, in historical occupational cohorts, for example, past exposure measurements made in the working environment are rarely available and therefore variables such as work assignment or membership in a union or professional society are generally used to classify individual exposure. These proxy variables are, at best, only crude markers of the true exposure levels and the available detail may be insufficient to address adequately specific research questions. It is, however, unlikely that the accuracy or completeness of these records would be different for those who developed the outcome of interest and those who did not, since the data were recorded before the individuals developed the outcome under study, and, in most cases, for reasons totally unrelated to the investigation. As long as exposure misclassification is independent of the outcome status (i.e., is non-differential), it will tend to dilute any true association between the exposure and the outcome (see Sections 2.7 and 13.1.2).

The historical approach can be particularly successful when biological specimens were stored in the past, so that up-to-date laboratory techniques can be used to measure past exposure. Access to serum banks, for example, permits measurement of exposure to infectious agents (as in Example 8.9) or chemical substances. This method minimizes inaccuracies in past exposure measurement, but the number of such biological specimen banks is limited and the stability of the biological marker during long periods of storage is often unknown.

Example 8.9. *Many studies have reported elevated titres of IgG antibody against capsid antigen of Epstein–Barr virus (EBV) and high prevalence of antibodies against early antigen in patients with Hodgkin's disease. However, the blood samples analysed had been collected after diagnosis and treatment for Hodgkin's disease. To evaluate whether enhanced activation of EBV preceded the development of Hodgkin's disease, a study was undertaken in collaboration with five serum banks located in the USA, Norway and Sweden, holding samples from over 240 000 persons. Patients who had subsequently been diagnosed with cancer were identified by linkage to hospital records and cancer registry records. Forty-three cases of Hodgkin's disease were identified and their stored serum samples were then tested for EBV (Mueller* et al.*, 1989).*

In many cohort studies, a single classification of exposure is made for each individual at the time of his/her entry into the study. This is appropriate for attributes that do not change with time. Frequently, however, changes in exposure levels for the factors of interest occur during the course of long-term follow-up. Individuals may change jobs, decide to stop smoking (as in Example 8.10), or adopt a low-fat diet. It may be possible to repeat the exposure measurements at intervals during the follow-up period, or information on changes may be available from historical records, allowing the risk of developing the disease to be studied in

relation both to the initial exposure status and to subsequent changes.

There may be other reasons for re-assessing the exposure status of the study subjects, particularly in long-term prospective studies. More refined methods of measuring the exposures of interest may become available in the course of the study or new scientific information about the disease may indicate the importance (or desirability) of measuring additional variables that were not measured initially.

Example 8.10. *In the British doctors study described in Example 8.3, the first questionnaire was sent to all registered British doctors in 1951. Further inquiries about changes in smoking habits were made in 1957, 1966, 1972, 1978 and 1990–91. The two last questionnaires also included additional questions on alcohol consumption and some other personal characteristics. To assess the extent of changes in smoking habits during the 40-year follow-up period, the smoking habits of the men who replied to the 1990–91 questionnaire were compared with the habits they reported in the initial questionnaire in 1951 (Table 8.2) (Doll* et al.*, 1994a).*

Table 8.2.
Smoking habits of male participants in the British doctors study who replied to the 1951 and 1990–91 questionnaires.[a]

Smoking habits, 1951	Smoking habits, 1990-91					
	Non-smoker	Former smoker	Current smoker			No. (%) in 1951
			Cigarette only	Cigarette and other	Pipe or cigar	
Non-smoker	2361	198	17	4	86	2666 (25)
Former smoker	0	1374	10	3	66	1453 (13)
Current smoker						
Cigarettes only	0	3355	535	47	446	4383 (41)
Cigarettes and other	0	897	74	31	308	1310 (12)
Pipe or cigar	0	695	16	2	287	1000 (9)
No. (%) in 1990–91	2361 (22)	6519 (60)	652 (6)	87 (1)	1193 (11)	10 812 (100)

[a] Data from Doll *et al.* (1994a).

In Example 8.10, there were marked changes in the smoking habits of the male British doctors during the 40-year follow-up period (Table 8.2). Sixty-two per cent of the male doctors reported to be current smokers in 1951. The corresponding figure in 1990–91 was only 18%. Such changes in exposure status can be taken into account in the analysis of cohort studies (see Section 8.7).

8.4 Measurement of potential confounding variables

Cohort studies are observational studies and therefore participants are not randomly allocated to the various exposure categories. This may lead to differences between the groups in terms of exposures other than the one being studied. This is of importance only if these other exposures are also risk factors for the particular disease (or other outcome) under study,

i.e., if these exposures are confounding variables. Thus, if we are studying an occupational exposure in relation to lung cancer, it is necessary to be sure that the 'exposed' and 'unexposed' groups have a similar smoking history. If they do not, *statistical adjustment* for differences in smoking must be made (see Chapters 13 and 14). In order to carry out this adjustment, data on smoking for each individual are required. These data must be as accurate as possible and of similar quality to the data on the exposure of primary interest.

In historical cohort studies, information on confounding factors is frequently missing. This is one of their main limitations. For instance, in many of the historical occupational cohorts set up to investigate the relationship between asbestos exposure and respiratory cancers, information on smoking habits was not available. In contrast, the collection of data on potential confounders can be built into the design of most prospective cohort studies, except when local or national rates are taken as the unexposed comparison group.

8.5 Measurement of outcome(s)

A major advantage of cohort studies is that it is possible to examine the effect of a particular exposure on *multiple outcomes*.

Many cohort studies make use of existing routine surveillance systems to ascertain the outcomes of interest. Such systems include cancer registries, death certification and other specialized surveillance systems. They allow tracing of study subjects and ascertainment of their outcomes at much lower cost than if it is necessary to personally contact the subjects. However, it is only possible to examine outcomes of the type which are recorded routinely by these systems and according to the way in which they are coded there. This is particularly important in studies that last for several decades, since major changes may be introduced during the study period in the way diseases are ascertained and coded by these surveillance systems (see Appendix A.2.2).

When no form of disease surveillance system exists, or when the outcome of interest is not routinely recorded by them, some form of surveillance of disease within the cohort has to be set up. For instance, the ascertainment of the outcomes of interest may be done through self-administered questionnaires sent regularly to all study subjects, through personal interviews, or by regular physical examination of all members of the cohort.

Regardless of the method chosen to ascertain the outcome(s) of interest, it is vital to ensure that it is used identically for subjects exposed and those not exposed. If possible, interviewers and any other persons involved in the ascertainment of the outcomes should be kept blind to the exposure status of the study subjects. Otherwise, there is potential to introduce measurement bias (see Section 13.1.2).

Cohort studies focus on disease development. In order for a disease to develop, it must, of course, be absent at the time of entry into the study. An initial examination of the potential study population may be required to identify and exclude existing cases of disease. Even so, it may still be impos-

sible to be absolutely certain that all individuals were disease-free at entry to the study, particularly for conditions with a long latent period (i.e., with a long interval from disease initiation to onset of clinical symptoms and signs). It is therefore usual to exclude disease events occurring during some time period immediately following entry into the study. For cancer, this is often the first 2–3 years of follow-up.

8.6 Follow-up

The *criteria for entry* into the cohort must be defined before the start of the study in a clear and unambiguous way. Individuals should enter the cohort, and contribute person-time at risk, only after all the entry criteria have been satisfied. In most cohort studies, participants will join the cohort at different points in time and therefore the exact date of entry of each subject should be recorded.

Methods must be set up at the start of the study to ensure adequate follow-up of the study subjects. In general, these involve periodic contacts with the individuals such as home visits, telephone calls or mailed questionnaires. Cohort studies of conditions which have a long induction period require follow-up of a very large number of subjects over many years. This is obviously a huge and costly enterprise. To minimize these difficulties, many cohorts are defined in terms of membership of a particular group (professional body, occupational pension plan, union, health insurance plan, college alumni), in which the study population can be more easily followed. Any routine surveillance system that exists may be used to trace and follow up the study subjects at much lower cost than if the investigators had to contact them personally.

The *criteria for exit* from the cohort should also be clearly defined. A date should be specified as the end of the follow-up period (at least for the current analysis). For instance, if death is the outcome of interest, the vital status on that date must be ascertained for all cohort members. All subjects whose vital status is known at that date should contribute person-time at risk until that date (or until their date of death if it occurred earlier). Those whose vital status is not known at that date should be considered as 'lost to follow-up' and the last date for which their vital status was known should be taken as the end of their contribution to person-time at risk.

It is essential that as high a proportion of people in the cohort as possible is followed up. Some people will migrate, some die and some change employment, but every effort should be made to ascertain their outcome(s). All of these factors may be influenced by the exposure and so incomplete follow-up may introduce selection bias (see Section 13.1.1).

8.7 Analysis

The first step in the analysis of a cohort study is to measure the incidence of disease (or of any other outcome of interest) in those exposed and in those unexposed and compare them.

(a)

		Exposure	
		Yes	**No**
Outcome	Yes	*a*	*b*
	No	*c*	*d*
Risk in exposed group (p_1) = $a/(a+c)$			
Risk in unexposed group (p_0) = $b/(b+d)$			
Risk ratio = p_1/p_0			
Risk difference[a] = $p_1 - p_0$			

(b)

	Exposure	
	Yes	**No**
Cases	*a*	*b*
Person-time at risk	y_1	y_0
Rate in exposed group (r_1) = a/y_1		
Rate in unexposed group (r_0) = b/y_0		
Rate ratio = r_1/r_0		
Rate difference[a] = $r_1 - r_0$		

[a] If the exposure is protective, the risk and rate differences should be calculated as $p_0 - p_1$ or $r_0 - r_1$, respectively (see Section 5.2.2)

Table 8.3.
Analysis of a cohort study (*a*) by risks; (*b*) by rates.

If all, or virtually all, cohort members were followed up for the same period of time, we can calculate risk as the measure of disease occurrence in each group (Table 8.3(*a*)). For example, if the period is uniformly five years, the five-year risk can be computed separately for the exposed and unexposed groups. Risk ratio and risk difference can then be calculated as measures of relative and absolute effect, respectively.

If study subjects have unequal follow-up periods, this must be taken into account in the analysis. Follow-up durations may differ markedly if subjects were recruited into the study population over a relatively long period of time, or if some were lost to follow-up during the course of the study (for example, by moving out of the area). One way of handling variable follow-up periods is to calculate rates which use person-years at risk (or person-months or person-days, etc.) as the denominator (Table 8.3(*b*)). With this approach, each subject contributes to the population at risk only as many years of observation as he/she is actually observed; thus if the subject leaves after one year, he/she contributes 1 person-year; if after 10, 10 person-years (see Section 4.2.2).

People may contribute person-years of observation to more than one subgroup. Suppose, for example, that in a five-year study, disease incidence is determined for each 10-year age subgroup. A person entering the cohort when he or she reaches the age of 48 years will contribute 2 person-years of observation to the 40–49 year-old subgroup and 3 person-years of observation to the 50–59 year-old subgroup (see Section 4.3.2). This may also happen with exposure categories if the study subjects change their exposure status over time. For instance, a person may be a smoker for a few years and then give up.

Example 8.11. *The Nurses' Health Study described in Example 8.5 is a cohort study of 121 700 US female registered nurses aged 30–55 years when the cohort was established in mid-1976. A total of 1799 newly diagnosed breast cancer cases were identified during the first 10 years of follow-up from mid-1976 to mid-1986. Analyses were then conducted to investigate the relationship between oral contraceptive use and risk of breast cancer. On the baseline questionnaire in mid-1976, the following question was asked: "If you are now using or have used oral contraceptives, please indicate intervals of oral contraceptive use starting from first use and continuing until the present time. If applicable, please indicate reasons for stopping". The same question was asked on subsequent biennial follow-up questionnaires.*

In response to the 1976 questionnaire, 7133 women reported that they were using oral contraceptives. Responses to the 1978, 1980, and 1982 questionnaires showed that 2399, 1168, and 302 women, respectively, were still using oral contraceptives. In 1984, none of the women were current users.

The information given in the 1976 questionnaire was used to classify nurses according to categories of oral contraceptive use ('non-users', 'past users' and 'current users') and each nurse started contributing person-time at risk to that category. Similarly, for each subsequent two-year interval, women contributed additional person-time of follow-up to each updated report of oral contraceptive use. The follow-up of women who developed breast cancer was truncated at the time their breast cancer was diagnosed (Romieu et al.*, 1989).*

In Example 8.11, women who changed their oral contraceptive status during the follow-up period would have contributed person-time at risk to different exposure categories. For instance, a woman who began using oral contraceptives at the start of 1978 and stopped by the end of 1984 would have contributed approximately 1.5 person-years to the non-user category (from the start of the study in mid-1976 to the end of 1977), 7 person-years to the current user category (from the start of 1978 to the end of 1984), and 1.5 person-years to the past user category (from the start of 1985 until the end of the follow-up in mid-1986). If that woman had developed breast cancer at the end of 1982, her person-time contribution would have been 1.5 person-years to the non-user category but only 5 person-years to the current user category (her person-time contribution would have been stopped at the time she developed breast cancer).

The outcomes of interest also need to be allocated to the different exposure categories. In our previous example, the breast cancer case should have been allocated to the current user category since it occurred during the time the woman was contributing person-years to this category. Once the person-time at risk and the outcomes are allocated to the relevant exposure categories, it is possible to estimate breast cancer incidence rates for each oral contraceptive use category by dividing the number of breast

cancer cases in each category by the corresponding total number of person-years (Example 8.12).

The results in Example 8.12 show that there was no statistically significant difference in the incidence of breast cancer between ever-users (past and current users were pooled in this analysis) and never-users of oral contraceptives.

Example 8.12. *In Example 8.11, the incidence of breast cancer among nurses aged 45–49 years at the time of their entry into the cohort was examined in relation to use of oral contraceptives.*

	Oral contraceptive use		Total
	Ever (current or past use)	Never	
Cases	204	240	444
Person-years at risk	94 029	128 528	222 557
Rate per 100 000 pyrs	217	187	199

[a] Data from Romieu *et al.* (1989)

Rate ratio = 217 per 100 000 pyrs/187 per 100 000 pyrs = 1.16

95% confidence for the rate ratio = 0.96 to 1.40

Rate difference = 217 per 100 000 pyrs – 187 per 100 000 pyrs = 30 per 100 000 pyrs

95% confidence interval for the rate difference = – 8 to 68 per 100 000 pyrs

$\chi^2 = 2.48$; 1 d.f.; P=0.12.

(Test statistics and confidence intervals were calculated using the formulae given in Appendix 6.1).

Table 8.4.
Distribution of breast cancer cases and person-years at risk among US female nurses aged 45–49 years at the time of their entry into the cohort according to oral contraceptive use.[a]

Most often the exposures we are interested in can be further classified into various levels of intensity. Smokers may be classified by number of cigarettes smoked per day, oral contraceptive users by total duration of use, and occupational exposures by intensity of exposure (often estimated indirectly from data on type of job or place of work in the factory) or duration of employment. If various levels of exposure are used in the cohort, we can examine *trends* of disease incidence by level of exposure. The conclusions from a study are strengthened if there is a trend of increasing risk (or decreasing, if the exposure is protective) with increasing level of exposure (i.e., if there is an *exposure–response* relationship).

In Example 8.13, non-users of oral contraceptives were taken as the unexposed baseline category. Rate ratios for each timing and duration category were calculated by dividing their respective rates by the rate of the baseline category. Thus, the rate ratio for current users who had used oral contraceptives for 48 or less months was calculated as 1220 per 100 000 pyrs/187 per 100 000 pyrs = 6.52 (95% confidence interval 2.43–17.53) (Table 8.5). This result suggests that risk might be raised among current

Example 8.13. *In Example 8.11, the risk of developing breast cancer was also examined according to timing and duration of oral contraceptive use (Table 8.5).*

Table 8.5.
Incidence of breast cancer among nurses aged 45–49 years at the time their entry into the cohort by timing and duration of oral contraceptive use.[a]

Timing Duration of use	Cases	Person-years	Rate (per 100 000 pyrs)	Rate ratio (95% confidence interval)
Non-users[b]	240	128 528	187	1.00
Current users				
≤ 48 months	4	328	1220	6.52 (2.43–17.53)
> 48 months	4	2263	177	0.95 (0.35–2.55)
				χ^2 test for trend = 0.46; $P = 0.50$
Past users[c]				
≤ 48 months	106	54 080	196	1.05 (0.84–1.32)
> 48 months	86	36 039	239	1.28 (1.00–1.64)
				χ^2 test for trend = 3.33; $P = 0.07$

[a] Data from Romieu *et al.* (1989).
[b] Taken as the baseline category.
[c] Information on duration is missing for four past users.
(The 95% confidence intervals and the χ^2 test for a linear trend in rates were calculated using the formulae given in Appendix 6.1.)

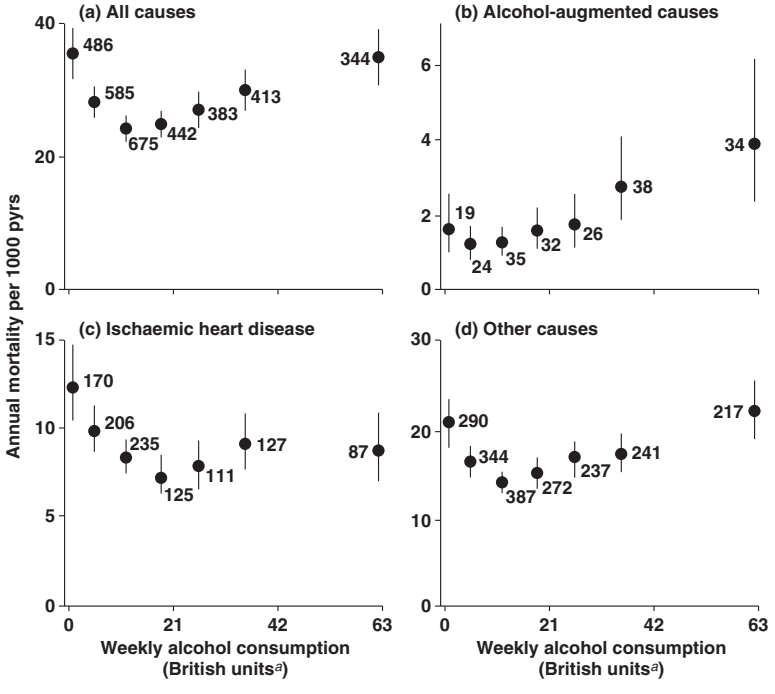
short-term users, but this estimate was based on only four breast cancer cases (hence, the wide confidence interval).

To assess whether there was a linear (positive or negative) trend in rates with increasing duration of use, a special statistical test (χ^2 test for a linear trend) was performed separately for current and past users (using the formula given in Section A6.1.3). There was moderate evidence of a positive trend among past users, but no evidence of a linear trend among current users (Table 8.5).

It should be noted that the shape of an exposure–response relationship does not have to be linear. This is clearly illustrated in Example 8.14. For instance, the relationship between alcohol consumption and all-cause mortality (Figure 8.2(*a*)) is U-shaped, with men who reported drinking 8–14 units of alcohol a week having the lowest mortality. By contrast, the relationship between alcohol consumption and mortality from alcohol-related disorders (Figure 8.2(*b*)) is basically linear, with a progressive increase in mortality with increasing alcohol consumption among regular drinkers. Thus, lack of a linear trend (as assessed by the χ^2 test for trend) does not imply absence of a relationship. The form of an exposure–outcome relationship should primarily be identified by plotting the data as in Figure 8.2. If the shape is suggestive of a non-linear trend, special statistical procedures, which are outside the scope of this book, should be used to assess its statistical significance.

Since the allocation of the study subjects to the different exposure categories is not random in cohort studies, the exposure groups are likely to differ in many respects other than the exposure of interest. These differences must be taken into account in the analysis. For instance, age is an

Example 8.14. *In the British doctors study described in Examples 8.3 and 8.10, additional questions on alcohol consumption were included in the 1978 questionnaire. Doctors were asked about frequency of drinking and, if they were regular drinkers (i.e., they drank in most weeks), about how much they drank in an average week. By 1991, almost a third of the 12 321 men who replied had died. The risk of death in men was then examined in relation to self-reported alcohol consumption (Doll* et al.*, 1994b). Some of the results are shown in Figure 8.2.*



[a] 1 unit = half pint (237 mL) of ordinary beer or one glass (125 mL) of wine or one glass (62 mL) of sherry or a glass (26 mL) of spirits.

Figure 8.2.
Male mortality from various causes by weekly alcohol consumption: (*a*) all causes; (*b*) conditions known to be alcohol-related (e.g., cancers of the liver, mouth, pharynx, oesophagus and larynx, cirrhosis, alcoholism, and external causes); (*c*) ischaemic heart disease; (*d*) other known causes (cerebrovascular diseases, respiratory diseases, all other cancers not included in (*b*), and others). Points and bars are rates and 95% confidence intervals adjusted for age, smoking habits and history of previous disease; the values are numbers of deaths (reproduced, by permission of BMJ Publishing Group, from Doll *et al.*, 1994b).

important confounding factor of the relationship between oral contraceptive use and breast cancer, since it is strongly associated with oral contraceptive use, and it is in itself an independent risk factor for breast cancer. Thus differences in the age distribution between women in different oral contraceptive categories may distort the relationship between oral contraceptive use and breast cancer incidence. To minimize this potential confounding effect, we deliberately restricted the analysis in Examples 8.12 and 8.13 to a narrow age-stratum (women aged 45–49 years at the time of their entry into the cohort). It is, however, possible (and desirable) to obtain summary measures (for all ages combined) that are 'adjusted' for age and any other potential confounding variable by using more complex

statistical methods. Standardization is one of these methods, as we shall see below. The interpretation of these 'adjusted summary measures' is similar, regardless of the method used. In Example 8.14 (and Figure 8.2), rates were adjusted for age and smoking habits. This means that differences in mortality between the different alcohol consumption categories cannot be explained by differences in their age or smoking distributions, provided the measurements of these confounding variables were valid. These issues are discussed further in Chapters 13 and 14.

Another common method of presenting the results of cohort studies, particularly those based on the disease experience of the general population as the comparison group, is to calculate *standardized mortality (or incidence) ratios* (see Section 4.3.3). Imagine that a total of 24 deaths from lung cancer were observed among a cohort of 17 800 male asbestos insulators. This observed number (O) is then compared with the number that would have been expected (E) if the cohort had the same age-specific mortality rates from lung cancer as the whole male population resident in the same area. Calculations similar to those shown in Section 4.3.3 indicate that only seven deaths would have been expected. Thus, the SMR is equal to $O/E = 24/7 = 3.4$ (or 340 if the SMR is expressed as a percentage). This measure is, in fact, an *age-adjusted rate ratio*. In this example, asbestos insulators were 3.4 times more likely to die from lung cancer than the entire male population resident in the same area, and this difference in mortality is not due to differences in the age-structure between the cohort and the general population. A similar approach was used in Example 8.15. Although this method is often used to adjust for age, it can also be used to adjust for any other confounding variable (e.g., calendar time, smoking habits). Statistical tests and 95% confidence intervals for an SMR can be calculated as shown in Sections A6.1.2 and A6.1.3.

Another way of analysing cohort data, which also takes into account different lengths of follow-up, is to use *survival analysis* methods. These are discussed in Chapter 12.

8.8 Cohort studies with nested case–control studies

In a traditional cohort study, all study individuals are subjected to the same procedures—interviews, health examinations, laboratory measurements, etc.—at the time of their entry into the study and throughout the follow-up period. Alternatively, a cohort may be identified and followed up until a sufficient number of cases develop. More detailed information is then collected and analysed, but only for the 'cases' and for a sample of the disease-free individuals ('controls'), not for all members of the cohort. This type of case–control study conducted within a fixed cohort is called a *nested case–control study* (see Chapter 9). This approach is particularly useful if complex and expensive procedures are being applied. For instance, blood samples for all members of the cohort can be collected at the time of entry and frozen. However, only the blood samples of

Example 8.15. *In the occupational cohort described in Example 8.7, the mortality experience of 6678 male rubber workers was compared with that of the 1968 US male population (McMichael et al., 1974). Mortality from selected causes of death is shown in Table 8.6.*

Table 8.6.
Mortality from selected causes of death among a cohort of male rubber workers.[a]

Cause of death (ICD-8 code)	Observed deaths (*O*)	Expected deaths (*E*)[b]	SMR (%) (100 × *O*/*E*)[c]	95% confidence interval
All causes	489	524.9	93	85–102
All neoplasms (140–239)	110	108.9	101	83–122
All malignant neoplasms (140–209)	108	107.3	100	82–120

[a] Data from McMichael *et al.*, 1974).

[b] Expected deaths calculated on the basis of the US male age-specific death rates, 1968.

[c] $P > 0.10$ for all the SMRs shown in the table.

(Test statistics and confidence intervals were calculated using the formulae given in Appendix 6.1.)

the cases (i.e., those individuals in the cohort who contract the disease under study) and of a subgroup of individuals who remained disease-free (the controls), are analysed at the end of the follow-up.

Example 8.16. *In 1972, a cohort of 42 000 children was established in the West Nile District of Uganda in order to investigate the etiological role of the Epstein–Barr virus (EBV) in Burkitt's lymphoma. A blood sample was obtained from each child at the time of entry into the study. By the end of the follow-up in 1979, 16 new Burkitt's lymphoma cases had been detected among the cohort members. The level of EBV antibodies in the serum sample taken at entry from each of these cases was then compared with the levels in the sera of four or five children of the same age and sex who were bled in the neighbourhood at the same time as the Burkitt's lymphoma case but who did not develop the disease ('controls') (Geser* et al.*, 1982).*

In Example 8.16, blood samples were obtained and stored from all 42 000 children who participated in the study but the rather complex and expensive virus tests were carried out only on the serum samples from the 16 children who developed the lymphoma and from a sample of about 80 selected disease-free members who acted as controls.

Similarly, in nutritional cohort studies, food diaries may be used to measure the subjects' usual dietary intake. As the coding and analysis of food diaries is very labour-intensive, a nested case–control study may be conducted in which only the diaries of the cases and of a sample of disease-free members of the cohort ('controls') are examined.

This type of study design and the analysis of its data are discussed further in Sections 9.3 and 9.5.

8.9 Interpretation

The main advantage of cohort studies is that the *time sequence* between exposure to a factor and occurrence of the disease is clearly observed.

Cohort studies may, however, suffer from important bias. Knowledge of the presence or absence of exposure to a particular factor may affect the subsequent assessment of disease and introduce *measurement bias*. This can occur if the decision as to the presence or absence of disease is made by persons who are aware of the subject's status with regard to the study factor. Cohort studies are also potentially prone to *selection bias* due to loss of study subjects. Such losses may occur initially if a portion of the target study population does not participate, or later on as members of the study population are lost to follow-up. These losses do not necessarily invalidate the study. However, the investigators should consider whether the reasons for loss might have affected the study results. Sometimes it is possible to gather information concerning lost subjects, particularly about whether they left because of illness or death possibly related to the exposures and outcomes under investigation.

As with any other observational study, *confounding* is a critical issue in the interpretation of cohort studies. Special statistical techniques can be used to take into account the effect of potential confounding variables, but only if these variables were known at the time of the data collection and if they were properly measured. If these data were not collected, we have to judge how much the observed findings are likely to have been affected by confounding in the light of all available biological and epidemiological evidence.

In Example 8.17, uranium miners had significantly elevated mortality for cancers at three sites relative to the general population (Table 8.7). Before concluding that these raised risks are due to exposures in the mines, we need to consider alternative explanations for the observed findings. A possible explanation for the high risk of lung cancer among miners is that they were heavier smokers than the general population. No information on smoking habits was available for the miners, but a survey of 697 men in other Czech uranium mines in 1974 showed that 76% were smokers, slightly more than the average (66%) for Czechoslovakian males at that time (see Tomášek *et al.* (1994)). The results from this survey indicate that differences in smoking habits are unlikely to have fully accounted for the estimated five-fold higher lung cancer risk in the miners than in the general population. Moreover, Table 8.8 (Example 8.18) shows that there was a significant positive trend of mortality from lung cancer with increasing cumulative exposure to radon. This trend provides considerable support for a true association between exposure to radon and lung cancer. Similar findings have been found in other radon-exposed miners.

Example 8.17. *A cohort of uranium miners in western Bohemia was identified by a retrospective search in 1970 of employment records. Workers were eligible for entry into the study if: (a) they started to work underground during 1948–59; (b) they worked there for at least four years; and (c) personnel and employment records were available. A total of 4320 miners were eligible. They were exposed to high radon levels, dust and, in one of the two major mines, also high levels of arsenic. The mortality experience of these miners up to the end of 1990 (an average of 25 years of follow-up) was then compared with that of the general population of the former Czechoslovakia. Information on the smoking and alcohol drinking habits of the miners was not available (Tomášek* et al.*, 1993, 1994). Mortality in this cohort from selected cancer sites is shown in Table 8.7.*

Cause of death	Observed deaths (*O*)	Expected deaths (*E*)[b]	SMR (*O*/*E*)[b] (95% confidence interval)
Lung cancer	704	138.6	5.08 (4.71–5.47)
Liver cancer	22	13.2	1.67 (1.04–2.52)
Cancer of gallbladder and extrahepatic bile ducts	12	5.3	2.26 (1.16–3.94)

[a] Data from Tomášek *et al.* (1993).

[b] Expected number of deaths (*E*) calculated using national male age-specific mortality rates for the former Czechoslovakia

Table 8.7.
Number of observed deaths (*O*) from selected cancer sites in the West Bohemian uranium miners cohort compared with the number expected (*E*) if the miners had the same mortality as the general male population of the former Czechoslovakia.[a]

Example 8.18. *In the cohort of uranium miners described in Example 8.17, the cumulative exposure to radon gas (and its progeny) was estimated for each miner. The exposure, measured in terms of 'working level months' (WLM), was calculated by considering the time spent in each mineshaft in conjunction with about 39 000 shaft-specific measurements of radon gas made in 1949–63 (Tomášek* et al.*, 1993). Mortality according to exposure levels is shown in Table 8.8.*

Site		Cumulative radon exposure (WLM) <110	110–149	150–209	210–329	≥330	*P*-value for trend
Lung cancer	*O*	86	100	139	161	181	<0.001
	O/*E* [b]	3.07	3.66	4.98	6.23	8.10	
Liver cancer	*O*	7	3	4	5	3	0.57
	O/*E* [b]	2.70	1.18	1.53	2.04	1.40	
Cancer of gallbladder and extrahepatic bile ducts	*O*	0	2	1	3	6	0.003
	O/*E* [b]	0.00	1.92	0.93	3.03	6.73	

[a] Data from Tomášek *et al.* (1993)

[b] Expected number of deaths (*E*) calculated using national male age-specific mortality rates for the former Czechoslovakia.

Table 8.8.
Number of observed deaths (*O*) and standardized mortality ratio (*O*/*E*) from selected cancers by cumulative radon exposure.[a]

There was also a positive trend in mortality from gallbladder and extrahepatic bile duct cancer with increasing levels of cumulative exposure to radon (Table 8.8), but there is little supporting evidence in favour of this finding from other epidemiological studies, and further investigation is needed. By contrast, mortality from liver cancer did not increase with cumulative radon exposure, making it unlikely that the excess in the miners was caused by radon. No information was available on the alcohol consumption of the miners, but they were well paid compared with other Czech workers and, therefore, it is likely that their alcohol consumption was higher than in the general population. They also had a significant excess of deaths from liver cirrhosis, probably caused by alcohol consumption and, for six of the liver cancer deaths, cirrhosis was also mentioned on the death certificate (Tomášek *et al.*, 1993).

The issues that need to be addressed in interpreting results from cohort studies are further discussed in Chapter 13.

Box 8.1. Key issues

- Cohort studies are studies in which subjects are selected on the basis of their exposure status and then followed up in time. In contrast with intervention studies, however, the allocation of exposure is not determined by the investigators.

- The main advantages of this type of study are:

    1. Exposure is measured before disease onset and is therefore likely to be unbiased in terms of disease development.

    2. Rare exposures can be examined by appropriate selection of study cohorts.

    3. Multiple outcomes (diseases) can be studied for any one exposure.

    4. Incidence of disease can be measured in the exposed and unexposed groups.

- The main disadvantages of this type of study are:

    1. They can be very expensive and time-consuming, particularly if conducted prospectively.

    2. Changes in exposure status and in diagnostic criteria over time can affect the classification of individuals according to exposure and disease status.

    3. Ascertainment of outcome may be influenced by knowledge of the subject's exposure status (*information bias*).

    4. Losses to follow-up may introduce *selection bias*.

- Cohort studies in cancer epidemiology generally involve the follow-up of a large number of individuals for long periods of time. They therefore tend to be very expensive and time-consuming. Various approaches may be used to reduce the costs:

    1. Use preexisting records (or biological specimens) to identify retrospectively a suitable study population and obtain information on the exposure status of their members *(historical cohort study*).

    2. Use available surveillance systems (e.g., death certification, cancer registration) to follow up subjects and obtain information on the outcomes of interest.

    3. Use national (or local) rates as the comparison unexposed group.

    4. Conduct a *nested case–control study.*

Further reading

* The book by Breslow & Day (1987) provides a very comprehensive coverage of the role, design, analysis and interpretation of cohort studies in cancer epidemiology. Some of the material is presented at a relatively advanced level.

* Discussion of the healthy worker effect can be found in papers by Carpenter (1987) and McMichael (1976).