1. THE ROLE OF COHORT STUDIES IN CANCER EPIDEMIOLOGY

1.1 Historical role

1.2 Present significance and specific strengths of cohort studies

1.3 Limitations of cohort studies

1.4 Implementation

1.5 Interpretation

1.6 Proportional mortality studies

CHAPTER 1

THE ROLE OF COHORT STUDIES IN CANCER EPIDEMIOLOGY

Longitudinal studies are of fundamental importance in human biology. In the study of physical growth, of mental and hormonal development, and in the process of ageing, the longitudinal approach has played a central role. The essential feature of such investigation is that changes over time are followed at the individual level. Most chronic diseases are the result of a process extending over decades, and many of the events occurring in this period play a substantial role. The longitudinal surveillance and recording of these events is therefore a natural model of study to obtain a complete picture of disease causation. Fortunately, for the study of a large number of chronic diseases, most of the relevant information on exposure can be summarized in a few relatively simple measures, so that continuous monitoring is not required. But the regular assessment of exposure variables may well be necessary, and in the epidemiology of cardiovascular disease, with its emphasis on physiological and biochemical explanatory measures, this approach has been the one of choice.

The essence of longitudinal studies in epidemiology is the identification of a group of individuals about whom certain exposure information is collected; the group is then followed forward in time to ascertain the occurrence of the diseases of interest, so that for each individual prior exposure information can be related to subsequent disease experience. Since the first requirement of such studies is the identification of the individuals forming the study group – or cohort – longitudinal studies in cancer epidemiology are usually referred to as *cohort* studies. (This use of the word 'cohort' first appeared in the literature in a demographic setting in 1944, according to the Oxford English Dictionary. It had apparently been introduced informally in 1935, as described by Wall & William, 1970.)

There are two ways in which the follow-up over time may be conducted. First, one may assemble the cohort in the present, and follow the individuals prospectively into the future. This type of study is often referred to as a *prospective cohort* study. It has the advantage that one may collect exactly the information thought to be required, and the disadvantage that many years may elapse before sufficient cases of disease have developed for analysis.

Second, one may identify a group with certain exposure characteristics, by means of historical records, at a certain defined time in the past, and then reconstruct the disease experience of the group between the defined time in the past and the present. This type of study has been called a *historical cohort* study. The advantage is that results are

potentially available immediately; the disadvantage is that the information available on the cohort may not be completely satisfactory, since it would almost certainly have been collected for other purposes. Much may be missing, and it may not correspond closely to the question of interest. The term 'retrospective cohort study' is also commonly used, but is slightly misleading, since the essential viewpoint in most such studies is forward in time, although starting in the past. The term 'historical cohort study' is preferable logically. In both types of study, the individuals comprising the cohort are identified, and information on their exposure obtained, before their disease experience is ascertained.

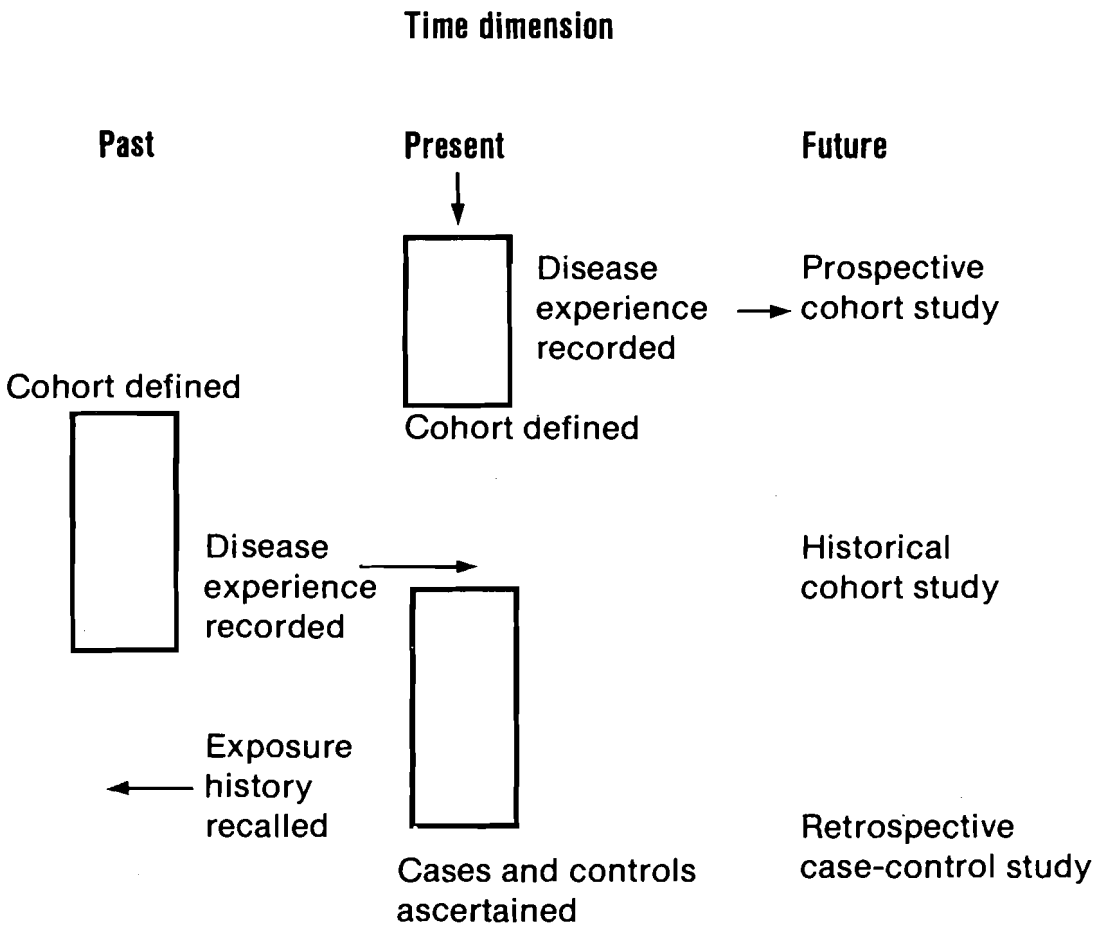
Cohort studies, by recording disease occurrence in a defined group, provide measures of incidence, or mortality rates, and it is these rates that provide the basic measures of disease risk. By allowing one to measure the basic risk associated with different levels and types of exposure, cohort studies provide the foundation of cancer epidemiology. It so happens, however, that a frequently convenient way of expressing the excess risk in one group compared to another is in terms of the ratio of the rates in the two groups, and to estimate the ratio of the rates one can use just a sample of the overall cohort. Since it is often easier and cheaper to obtain information on a sample rather than on the entire cohort, the case-control study has become widely adopted in cancer epidemiology as an alternative to the cohort study.

In fact, as commonly used, the case-control approach departs more radically from a cohort study than simply by sampling. In many case-control studies, the individuals with the disease in question and some comparison group are ascertained first, and their exposure experiences for some defined period of time in the past obtained retrospectively. The results are used to derive rate ratios. A cohort study faces forwards in time, starting with the defined population and its exposure status, and observing the subsequent disease experience, whereas a retrospective case-control study faces backwards in time, starting with the disease status and reconstructing the exposure history from which it emerged. Graphically, the distinction can be expressed as shown in Figure 1.1

Notwithstanding these differences, however, the rate ratios estimated in a case-control study should refer to rates in some defined population. As argued in Volume 1 of this series, the inferences one draws from the results of a case-control study depend logically on the interpretation one can give to it as having arisen by sampling from some underlying cohort. The less clear the definition of the underlying population, the less confidence can be put in the results of the case-control study. Thus, although the case-control and cohort approaches appear clearly distinct, they share the same logical framework of inference. An increasing number of studies have components of both approaches in their design. In these hybrid designs, the cohort component would usually identify the group and ascertain the disease experiences in the follow-up period; the exposure experience would then be obtained using the case-control approach. In this way, one ensures strict definition of the study cohort, but the effort and resources devoted to obtaining accurate exposure data can be concentrated on the most informative individuals. We discuss later at some length (§1.4*i*) the interplay between the cohort and the case-control approach.

Common to both cohort and case-control studies is the extended period of

Fig. 1.1 Differences between cohort and case-control studies



observation, relating to disease experience in the former and to exposure experience in the latter, and sometimes both in either case, and the fact that the individual is the unit of observation. These two features contrast with those of studies in which populations are compared by using cross-sectional data on both exposure and disease occurrence – so-called 'population correlation' or 'ecological' studies. This type of study would normally be given little weight in assessing the basic causality of a relationship, and, in the series of *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans,* a prerequisite for evidence to be deemed sufficient to establish carcinogenicity in humans is that it derive from individual-based studies. Correlation studies may be useful in suggesting interesting areas of study, that is, for hypothesis generation. The distinctions, however, are not absolute. Population comparisons may be made on the basis of temporal changes or of the experience with respect to exposure and disease of different birth cohorts, rather than among populations defined geographically, and such comparisons are often given greater weight. A cohort study, on the other hand, may include little or no information on variations in exposure between individuals, it being known simply that the cohort as a whole was exposed – for example, had received Bacillus Calmette–Guerin (BCG) vaccination in the first year of life.

1.1 Historical role

In 1954, two papers were published that are landmarks in the historical development of cancer epidemiology. The first, called a 'preliminary report', described the rationale for, and the first results of, the prospective cohort study of British doctors (Doll & Hill, 1954), designed to investigate the relatonship of tobacco smoking to lung cancer. The second, a historical cohort study, reported on the risk of bladder cancer in the British chemical industry (Case *et al.*, 1954; Case & Pearson, 1954).

The prospective study of British doctors was initiated in 1951, when the results of a number of case-control studies had already been published demonstrating an association between lung cancer and cigarette smoking. (The design and execution of the study are described in detail in Appendix IA.) It is interesting to examine why, in view of the results of the case-control studies, a large scale, long-term study was felt necessary. The 1954 paper by Doll and Hill starts as follows:

'In the last five years a number of studies have been made of the smoking habits of patients with and without lung cancer. All these studies agree in showing that there are more heavy smokers and fewer nonsmokers among patients with lung cancer than among patients with other diseases. While, therefore, the various authors have all shown that there is an "association" between lung cancer and the amount of tobacco smoked, they have differed in their interpretation. Some have considered that the only reasonable explanation is that smoking is a factor in the production of the disease; others have not been prepared to deduce causation and have left the association unexplained.

'Further retrospective studies of that same kind would seem to us unlikely to advance our knowledge materially or to throw any new light upon the nature of the association. If, too, there were any undetected flaw in the evidence that such studies have produced, it would be exposed only by some entirely new approach. That approach we considered should be "prospective". It should determine the frequency with which the disease appeared, in the future, among groups of persons whose smoking habits were already known.'

In this initial report on the British doctors study, the authors stressed that the results of the prospective study were in close agreement (Table 1.1) with the results of their earlier case-control study (Doll & Hill, 1950), in terms of the ratios of the rates in the different smoking categories. The absolute level of the rates, however, appeared to be more than twice as high in the case-control study (confined to the subset of the study consisting of residents of Greater London) than in the cohort of doctors. It should be noted that the results of the case-control study were converted into absolute incidence rates for lung cancer and were not limited to a description of the effect of smoking in terms of the ratios of rates in the different smoking categories.

The results of 20 years or more of follow-up have been published in some detail (Doll & Peto, 1976, 1978; Doll *et al.*, 1980). A comparison of these results with those of the case-control study published in the early 1950s (Doll & Hill, 1950, 1952) highlights the relative merits of the two approaches. The case-control study was begun in April 1948, and the final results published in December 1952. A total of 4342 people were interviewed, of whom 1488 were lung cancer cases. Most of the analyses referred

Table 1.1 Comparison of the relation between risk of dying from lung cancer and the most recent number of cigarettes smoked per day, among men aged 45–74, obtained from a prospective cohort study and a retrospective case-control study

	Non-smokers	Smokers			All groups
		1–14 cig./day	15–24 cig./day	25+ cig./day	
Standardized rates:					
'Backward' study[a] of patients' histories	0.11	1.56	2.20	4.00	1.97
'Forward' study[b] of mortality of doctors	0.00	0.50	0.97	1.45	0.73
Each rate as a % of the rate for all groups:					
'Backward' study of patients' histories	6%	79%	112%	203%	100%
'forward' study of mortality of doctors	0%	68%	133%	199%	100%

[a] From Doll and Hill (1950)
[b] From Doll and Hill (1954)

to 1465 lung cancer cases and a series of 1465 individually matched controls. By contrast, the prospective study was begun in October 1951, the month the British doctors were first approached, and the most recent results for men, based on 20 years of follow-up, appeared in 1978, and for women, based on 22 years of follow-up, in 1980. During these years, 441 lung cancer deaths were registered among the 34 440 men, and 27 among the 6194 women, enrolled in the study. The advantages of the case-control study are clear: many more cases of lung cancer could be assembled in a much shorter time. In addition, the total number of persons interviewed in the case-control study was only one-tenth the number who completed the questionnaire in the prospective study. This reduction in numbers facilitates the asking of a broader range of questions, allowing one to obtain information on a wider range of potential risk factors. In the prospective study, the questionnaire was kept short and simple, in order, as the authors say, 'to encourage a high proportion of replies'.

What was achieved in return, then, for the high cost and length of the prospective study? Part of the answer is given by comparing Table 1.2a, from the prospective study, and Table 1.2b, from the case-control study. Attention has been limited to males; a similar comparison could be made for females. The results of the case-control study with regard to the health effects of cigarette smoking, relative to the average amount smoked per day, are summarized in Table 1.2b. For the prospective study, in addition to the 441 lung cancer deaths, there were 9631 other deaths, and the full range of the effect of cigarette smoking on mortality can be examined, either for each individual cause of death or for all causes combined. One can see that there is nearly a two-fold difference in the annual death rate between heavy smokers and nonsmokers.

Table 1.2a Death rates between November 1951 and October 1971 by cause of death and by smoking habits when last asked: male British doctors[a]

Cause of death	No. of deaths	Annual death rate per 100 000 men, standardized for age							χ^2	
		Non-smokers	Current or ex-smokers	Ex-smokers	Current smokers, any tobacco	Current smokers, any tobacco (cig./day)			Others *versus* non-smokers[b]	Trend[b]
						1–14	15–24	≥25		
Cancer										
Lung	441	10	83	43	104	52	106	224	41.98	197.04
Oesophagus	65	3	12	5	16	12	13	30	3.94	14.94
Other respiratory sites	46	1	9	4	11	6	9	27	3.31	21.68
Stomach	163	23	28	21	32	28	38	32	—	—
Colon	195	27	34	34	34	35	33	31	—	—
Rectum	78	6	14	14	14	10	14	27	2.81	10.76
Pancreas	93	14	16	12	18	14	18	27	—	3.98
Prostate	186	39	30	31	30	28	31	38	—	—
Kidney	46	3	8	9	8	8	9	9	—	—
Bladder	80	9	14	11	16	16	16	12	—	—
Marrow and reticulo-endothelial system	152	33	24	26	24	27	22	19	—	(3.51)
Unknown site	64	12	11	9	12	10	13	14	—	—
Other site	151	25	26	29	24	19	24	35	—	—
Respiratory disease										
Respiratory tuberculosis	57	3	11	11	10	8	7	21	3.83	10.51
Asthma	40	4	7	12	5	5	7	0	—	—
Pneumonia	345	54	59	62	57	47	62	91	—	6.94
Chronic bronchitis and emphysema	254	3	48	44	50	38	50	88	25.58	47.23
Other respiratory disease	121	16	21	24	19	20	14	26	—	—
Pulmonary heart disease	50	0	9	7	11	9	10	19	4.72	8.37
Cardiac and vascular disease										
Rheumatic heart disease	77	14	13	12	13	14	16	5	—	—
Ischaemic heart disease	3191	413	554	533	565	501	598	677	22.59	53.56
Myocardial degeneration	615	67	108	98	116	111	111	160	9.58	13.92
Hypertension	239	37	41	41	41	33	43	58	—	4.67
Arteriosclerosis	117	21	20	17	21	17	21	46	—	4.85
Aortic aneurysm (non-syphilitic)	121	5	22	16	26	18	28	45	8.40	25.60
Venous thromboembolism	48	9	8	8	8	8	5	14	—	—
Cerebral thrombosis	616	86	106	105	107	92	123	131	—	0.54
Other cerebrovascular disease	692	107	118	122	115	112	114	128	—	—
Other cardiovascular disease	267	53	44	49	41	37	42	52	—	—

[a] From Doll and Peto (1976)

[b] Figures are given whenever the value was greater than 2.71 ($p<0.1$); figures in parentheses indicate a decreasing trend from nonsmokers to heavy smokers; others indicate an increasing trend

Table 1.2a—*contd.*

Cause of death	No. of deaths	Annual death rate per 100 000 men, standardized for age							χ^2	
		Non-smokers	Current or ex-smokers	Ex-smokers	Current smokers, any tobacco	Current smokers, any tobacco (cig./day)			Others *versus* non-smokers[b]	Trend[b]
						1–14	15–24	≥25		
Other diseases										
Parkinsonism	51	14	8	13	5	8	1	4	—	(9.10)
Peptic ulcer	79	8	14	12	15	10	20	23	—	8.26
Cirrhosis of liver, alcoholism	80	7	14	10	16	10	10	40	—	22.53
Hernia	16	0	3	2	4	3	4	7	—	4.16
Other digestive disease	144	20	25	27	24	18	33	26	—	3.25
Nephritis	79	10	14	10	16	15	14	21	—	—
Other genitourinary disease	136	19	23	24	23	22	24	26	—	—
Other disease	391	59	67	73	64	65	58	73	—	—
Violence										
Suicide	173	21	31	27	32	30	28	46	—	6.26
Poisoning	74	9	13	6	16	16	14	26	—	6.86
Trauma	240	46	39	36	41	47	25	56	—	—
All causes (no. of deaths)	10 072	1317 (490)	1748 (9132)	1652 (3114)	1802 (6018)	1581 (2707)	1829 (1986)	2452 (1325)	68.47	244.16

Table 1.2b Most recent amount of tobacco smoked regularly before the onset of the present illness: lung carcinoma patients and matched control patients with other diseases (males only)[a]

Disease group	Number of non-smokers	Number smoking daily:				
		1 cig.	5 cig.	15 cig.	25 cig.	50 cig.
1357 lung cancer patients	7	49	516	445	299	41
	0.5%	3.6%	38.0%	32.8%	22.0%	3.0%
1357 control patients	61	91	615	408	162	20
	4.5%	6.7%	45.3%	30.1%	11.9%	1.5%

[a] From Doll and Hill (1952)

For an exposure with a wide range of deleterious effects, there is no substitute for the broad picture given by Table 1.2a.

A second advantage to be gained from the extended duration of a prospective study is the opportunity it affords to obtain further information on the exposure of interest. In the British doctors study, four separate questionnaires were sent (in 1951, 1957, 1963 and 1971). The good compliance of the population under study is well indicated by the low proportion of non-responders to the second, third and fourth questionnaires

Table 1.3 Response to questionnaires

	Second questionnaire	Third questionnaire	Fourth questionnaire
Survey period	November 1957–October 1958	March–October 1966	July–October 1972
No. known to have died before end of survey period	3122	7301	10 634
No. presumably alive at end of survey period	31 318	27 139	23 806
No. who replied by end of survey (and % of men then alive)	30 810 (98.4)	26 163 (96.4)	23 299 (97.9)
Reasons for non-response:			
Too ill	31	65	21
Refused	36	63[a]	102[a]
Address not found	72	403	22
Unknown and other reasons	369	445	362

[a] Includes all men who had refused previously

(see Table 1.3); it was not sent to those who had refused to reply previously or who had been struck off the Medical Register. These additional questionnaires certainly improved the quality of the basic information that was being sought, namely, the average amount smoked in the few years preceding onset of disease, and also provide much useful information on the time sequence of events, particularly changing smoking habits. The relationship between the years since stopping smoking and the level of excess risk for lung cancer, both absolute and relative, has been more clearly defined from the prospective studies.

The British doctors prospective study was followed rapidly by a similar study undertaken by the American Cancer Society, started in 1952 (Hammond, 1966), and two years later, in 1954, by a study of United States veterans (Kahn, 1966). Other studies have followed since, notably a prospective study in Japan (Hirayama, 1975). The impact of these studies was much greater than their unambiguous demonstration of the health effects of tobacco smoking. They were the studies which, at least in the field of cancer, established chronic disease epidemiology as a rigorous scientific discipline.

The case-control studies, when they were first reported, appeared fraught with possible biases. The potential for error, so many claimed, was such that little credence could be put in the results. The large prospective studies begun in the early 1950s have shown that observational studies in humans can produce results that establish beyond reasonable doubt associations between exposure and disease. Furthermore, they demonstrated pragmatically that prospective cohort studies and retrospective case-control studies can, under favourable circumstances, give the same results. This demonstration, complementing the theoretical arguments developed at that time for the equivalence of the two study designs, at least in terms of estimating relative risks

Table 1.4 The number of death certificates expected if no special risk were operating and the number of cases and death certificates found for the various exposure classes[a]

Rank	Class	Group	Total no. of cases found	No. of cases on nominal roll			No of cases on nominal roll for whom death certificate mentions bladder tumour	Expected no. of such cases	% of expected no. derived from incomplete data	Significance of difference	*P*
				Total	Alive	Dead					
1	Aniline without magenta contact	I	4	4[b]	2[b]	2[b]	1	0.30	35.8	None	>0.1
		II	0	0	0	0	0	0.23		None	>0.1
		III	0	0	0	0	0	0.01		None	>0.1
		All	4	4[b]	2[b]	2[b]	1	0.54		None	>0.1
2	Aniline with possible magenta contact	I	8	5	3	2	2	0.30	15.6	Suspicious	0.025
		II	1	1	0	1	1	0.05		None	>0.1
		III	0	0	0	0	0	0.00		None	>0.9
		All	9	6	3	3	3	0.35		Signicant	<0.02
3	All aniline	I	12	9[b]	5[b]	4[b]	3	0.60	20.3	Suspicious	0.025
		II	1	1	0	1	1	0.28		None	>0.01
		III	0	0	0	0	0	0.01		None	>0.1
		All	13	10[b]	5[b]	5[b]	4	0.89		Suspicious	0.025
4	Benzidine	I	38	34	21	13	10	0.54	3.7	Very high	<0.001
		II	0	0	0	0	0	0.17		None	>0.1
		III	0	0	0	0	0	0.01		None	>0.1
		All	38	34	21	13	10	0.72		Very high	<0.001
5	α-Naphthylamine	I	28	19	13	6	6	0.66	3.2	High	0.005
		II	0	0	0	0	0	0.04		None	>0.1
		III	0	0	0	0	0	0.00		None	>0.9
		All	28	19	13	6	6	0.70		High	<0.005
6	β-Naphthylamine	I	59	55	28	27	26	0.30	4.1	Very high	<0.001
		II	0	0	0	0	0	0.00		None	>0.9
		III	0	0	0	0	0	0.00		None	>0.9
		All	59	55	28	27	26	0.30		Very high	<0.001
7	Mixed exposures	I	162	135	50	85	75	1.15	13.5	Very high	<0.001
		II	9	7	0	7	5	0.32		High	<0.005
		III	2	2	1	1	1	0.006		Significant	<0.005
		All	173	144	51	93	81	1.48		Very high	<0.001
8	All classes, excluding aniline	I	287	243	112	131	117	2.65	7.3	Very high	<0.001
		II	9	7	0	7	5	0.53		High	<0.005
		III	2	2	1	1	1	0.02		Suspicious	0.025
		All	298	252	113	139	123	3.20		Very high	<0.001
9	All classes	I	299	252	117	135	120	3.25	9.3	Very high	<0.001
		II	10	8	0	8	6	0.81		High	<0.005
		III	2	2	1	1	1	0.03		Suspicious	0.025
		All	311	262	118	144	127	4.09		Very high	<0.001

[a] From Case *et al.* (1954)
[b] Also manufacturer of auramine

(Cornfield, 1951), has led to the case-control study becoming the major methodological tool in cancer epidemiology.

The bladder cancer study of the British chemical industry (Case *et al.*, 1954) also played a seminal role in the evolution of cancer epidemiology and is the prototype of historical cohort studies. Its purpose was to determine 'whether the manufacture or use of aniline, benzidine, β-naphthylamine or α-naphthylamine could be shown to produce tumours of the urinary bladder in men so engaged'. It had been suspected since the last century that the production of aniline-based dyestuffs might produce bladder cancers among the men employed. There was lack of unanimity concerning the agent or agents responsible, however, and little information on the level of the excess risk.

In the early 1950s, Case and his co-workers constructed a list, or nominal roll as they termed it, of all those who had ever been employed in the chemical industry in the United Kingdom for at least six months since 1920, worked for one of the 21 firms which cooperated in the study, and for whom exposure to one of these compounds listed above had been documented. Age and the dates between which exposure to these substances occurred were recorded. A search was made retrospectively for all bladder cancer cases occurring among men who had been employed, in or after 1921 until 1 February 1952, in the chemical industry. Of the 455 cases identified, 127 were on the nominal roll, had died, and had bladder cancer mentioned on the death certificate. Since bladder cancer death rates based on death certificates mentioning bladder cancer were known between 1921 and 1952, the nominal roll could be used to calculate expected numbers, strictly comparable to the 127 observed bladder cancer deaths, using calendar time- and age-specific rates. The results of these calculations are given in Table 1.4. Accepting the authors' use of the terms 'aniline', 'benzidine' and so on to mean these substances as encountered in industrial practice, rather than to mean the pure chemicals, Table 1.4 gives clear, quantitative evidence of the carcinogenicity to humans of β-naphthylamine, α-naphthylamine and benzidine. Aniline exposure, as it occurred in the British chemical industry in the first half of this century, presents a risk to the human bladder of a lower order of magnitude than the risk associated with β-naphthylamine, if it presents a risk at all. It is interesting to note that, 28 years later, in 1982 (IARC, 1982a), this study was still considered the soundest evidence on which to base an evaluation of the carcinogenicity of aniline to humans.

No real alternative existed to the strategy adopted by Case, since an answer to the question was urgently required. A prospective study was therefore out of the question, and, furthermore, exposure had already been substantially reduced so that present levels were no longer indicative of past exposure. Since only a very small proportion of all bladder cancer cases in the general population of England and Wales were related to the chemical industry, a general case-control study of bladder cancer would not have been informative. Reconstruction of the past for a cohort of individuals with recorded exposure to the compounds of interest was the only feasible approach. In the 30 years since Case's study was reported, this methodology has become the approach of choice in many situations.

1.2 Present significance and specific strengths of cohort studies

In the next two sections, we discuss the relative merits and drawbacks of cohort and case-control studies. Although, as we have seen, the distinction is not always clear cut,

and the two may merge into each other, what we have in mind in the following discussion is a comparison of two approaches: one in which a group of individuals is defined, their exposure determined and their subsequent disease experience ascertained; the other in which cases of a specific disease are identified together with a suitable comparison group, and information on exposure before disease onset obtained retrospectively. Described in this way, it would seem natural that the latter might appeal if the focus is on causation of a specific disease, and the former if interest is on the health consequences of a given exposure.

Certainly, cohort studies have played a major role in the last 30 years in identifying specific environmental agents or other factors as carcinogenic hazards. We give in Table 1.5 the factors that are currently recognized as causally related to cancer risk in man, together with the type of evidence on which causality has been established. Case reports have been excluded. The intention has been to categorize the first epidemiological study that could be regarded as conclusive, although the choice is necessarily subjective, at least on occasion. For some associations, such as that between sexual activity and cervical cancer, the first epidemiological study establishing the link is not readily identifiable. A series of studies over the years has refined the nature of the association. For others, the effect is so strong that a case series, complemented by theoretical calculation of the size of the expected number, has been sufficient to establish the existence of an excess. The induction of lymphomas following immunosuppression of recipients of renal grafts using azathioprine is an example, but the excess of other malignancies emerged only from a formal cohort study. An immediately evident feature of Table 1.5 is that the cohort study has been the method used to incriminate the great majority of factors so far identified as carcinogenic hazards. In addition to their value in establishing qualitatively that a carcinogenic hazard exists, cohort studies have been of importance in establishing quantitative estimates of increased risk. In Table 1.6 we list the few agents given in Table 1.5 for which substantial quantitative information is available on dose-response or the temporal evolution of risk. In later chapters, particularly Chapter 6, we discuss the quantification of excess risk and its temporal evolution in considerable detail, but one can see from Table 1.6 that much of the information currently available, particularly on the temporal development of risk, has come from cohort studies.

Tables 1.5 and 1.6 outline the significance that cohort studies have had historically in cancer epidemiology. These tables might give the impression that cohort studies are mainly of value when studying specific exposures, often rare and of little relevance to the great majority of cancers. Certainly, for the factors identified as cancer risks to which exposure is widespread, case-control studies have often been the study method used. The present significance of cohort studies, however, is wider than that suggested by Table 1.5.

Although, in many situations, the relatively low cost of retrospective case-control studies and the speed with which they can be conducted make them the design of choice, there are clearly occasions in which such an approach is inadequate, and a study design is required that is directed more towards the continuous recording of events in the years before disease, or which focuses on a broad spectrum of disease. This approach is the essence of a longitudinal or cohort study, the strengths of which

Table 1.5 Established human carcinogenic agents and circumstances

Agent	Site affected	Type of exposure	Main type of evidence[a]
Aflatoxin	Liver	Food	Geographic[1]
Alcoholic drinks	Mouth, Larynx, Oesophagus, Pharynx	Lifestyle	Case-control[2,3,4,5]
4-Aminobiphenyl	Bladder	Occupational	Cohort[6]
Analgesic mixtures containing phenacetin	Renal pelvis	Medicinal	Case-control[7]
Arsenic and arsenic compounds	Skin	Medicinal, occupational, drinking-water	? Case series Geographical[8]
	Lung	Occupational	Cohort[9]
Asbestos	Lung, Pleura, Peritoneum	Occupational and geographic	Cohort[10,11] (in an informal sense)
Auramine manufacture	Bladder	Occupational	Cohort[12]
Azathioprine	Lymphomas Squamous skin tumours Liver	Medicinal	Cohort[13,14] (after earlier reports of a very high incidence of lymphomas)
Benzene	Leukaemia	Occupational	Cohort[15]
Benzidine	Bladder	Occupational	Cohort[16]
Betel-quid and tobacco chewing	Oral cavity, Oesophagus	Lifestyle	Case-control[17]
Bis-chloromethyl ether	Lung	Occupational	Cohort[18,19]
Boot and shoe (leather goods) manufacture	Nasal sinus	Occupational	Based on cases, but interpreted mainly as a case-control study[20]
Busulphan (myleran)	Leukaemia	Medicinal	Cohort[21]
Chlorambucil	Leukaemia	Medicinal	Cohort[22]
Chlornaphazine	Bladder	Medicinal	Cohort[23]
Chromium and certain chromium compounds	Lung	Occupational	Cohort[24]
Conjugated oestrogens	Endometrium	Medicinal	Case-control[25,26]
Cyclophosphamide	Bladder	Medicinal	Cohort[13,14]
	Leukaemia	Medicinal	Case-control within a cohort[27]
Diethylstilboestrol	Vagina	Medicinal	Case-control[28]
Furniture manufacture	Nasal sinus	Occupational	Cohort[29]
Ionizing radiation	Leukaemia	Occupational, medicinal	Cohort[30,31]
	Most other sites	Warfare	Cohort[32] (see references to Appendix IB)
Isopropyl alcohol manufacture	Nasal sinus	Occupation	Cohort[33]
Melphalan	Leukaemia	Medicinal	Cohort[34]
Methoxsalen with UV-A (PUVA)	Skin	Medicinal	Cohort[35] Case-control within a cohort[35]

Table 1.5 (*contd*)

Agent	Site affected	Type of exposure	Main type of evidence[a]
Mustard gas	Lung, larynx	Occupational	Cohort[36]
β-Naphthylamine	Bladder	Occupational	Cohort[16]
Nickel refining	Nasal sinus Lung	Occupational	Cohort[37,38]
Obesity	Endometrium Gallbladder	Lifestyle	Case-control[39] Cohort[40]
Sexual promiscuity	Cervix	Lifestyle	Case-control[41]
Soots, tars and oils	Scrotum Skin Lung	Occupational	Numerous industrial cohorts
Tobacco smoking	Many sites	Lifestyle	Both case-control and cohort (see § 1.1)
Treosulphan	Leukaemia	Medicinal	Cohort[42]
Ultraviolet light	Skin	Lifestyle (occupational)	Geographic and other[b,43]
Vinyl chloride	Liver (angiosarcoma) (lung, brain)	Occupational	Based on cases[44] but interpreted as a cohort study[45]
Hepatitis B virus	Liver	Lifestyle	Cohort (see Appendix IC)
Reproductive history, age at first birth, age at menarche, age at menopause, parity	Breast	Lifestyle	Case-control[46]
	Ovary		Case-control[47]
Chlonarchis siensis	Liver (cholangio-carcinoma)	Lifestyle	Geographic[48]
Schistosoma haematobium	Bladder	Lifestyle	Geographic[49]
Epstein-Barr virus	Burkitt's lymphoma	Lifestyle	Cohort[50]

[a] References: [1] Peers, F.G. & Linsell, C.A. (1973) Dietary aflatoxin and liver cancer. A population based study in Kenya. *Br. J. Cancer*, **27**, 473–484; [2] Wynder, E.L., Bross, I.D.J. & Feldman, R.M. (1957) A study of the etiological factors in cancer of the mouth. *Cancer*, **10**, 1300–1323; [3] Wynder, E.L., Bross, I.J. & Day, E. (1956) A study of environmental factors in cancer of the larynx. *Cancer*, **9**, 86–110; [4] Wynder, E.L. & Bross, I.J. (1961) A study of etiological factors in cancer of the esophagus. *Cancer*, **14**, 389–413; [5] Vogler, W.R., Lloyd, J.W. & Milmore, B.K. (1962) A retrospective study of etiological factors in cancer of the mouth, pharynx and larynx. *Cancer*, **15**, 246–258; [6] Melick, W.F., Naryka, J.J. & Kelly, R.E. (1971) Bladder cancer due to exposure to para-aminobiphenyl: a 17-year follow-up. *J. Urol.*, **106**, 220–226; [7] McCredie, M., Ford, J.M., Taylor, J.S. & Stewart, J.H. (1982) Cancer of the renal pelvis in New South Wales. Relationship to analgesic consumption and smoking. *Cancer*, **49**, 2617–2625; [8] Tseng, W.P., Chu, H.M., How, S.W., Fong, J.M., Lin, C.S. & Yeh, S. (1968) Prevalence of skin cancer in an endemic area of chronic arsenicism in Taiwan. *J. natl Cancer Inst.*, **40**, 453–463; [9] Lee, A.M. & Fraumeni, J.F., Jr (1969) Arsenic and respiratory cancer in man: an occupational study. *J. natl Cancer Inst.*, **42**, 1045–1052; [10] Doll, R. (1955) Mortality from lung cancer in abestos workers. *Br. J. ind. Med.*, **12**, 81–86; [11] Wagner, J.C., Sleggs, C.A. & Marchand, P. (1960) Diffuse pleural mesothelioma and asbestos exposure in the North-western Cape Province. *Br. J. ind. Med.*, **17**, 260–271; [12] Case, R.A.M. & Pearson, J.T. (1954) Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. Part II. Further considerations of the role of aniline and of the manufacture of auramine and magenta (fuchsine) as possible causative agents. *Br. J. ind. Med.*, **11**, 213–216; [13] Kinlen, L.J., Sheil, A.G.R. & Peto, J. (1979) Collaborative United Kingdom–Australasian study of cancer in patients treated with immunosuppressive drugs. *Br. med. J.*, **iv**, 1461–1466; [14] Kinlen, L.J., Peto, J., Doll, R. & Sheil, A.G.R. (1981) Cancer in patients treated with immunosuppressive drugs. *Br. med. J.*, **i**, 474; [15] Infante, P.F., Rinsky, R.A., Wagoner, J.K. & Young, R.J. (1977) Leukaemia in benzene workers. *Lancet*, **ii**, 868–869; [16] Case, R.A.M., Hosker, M.E., McDonald, D.B. & Pearson, J.T. (1954) Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. Part I. The role of aniline, benzidine, alpha-naphthylamine and beta-naphthylamine. *Br. J. ind. Med.*, **11**, 75–104; [17] Sanghvi, L.D., Rao, K.C.M. & Khanolkar, V.R. (1955) Smoking and chewing of tobacco in relation to cancer of the upper alimentary tract. *Br. med. J.*, **i**, 1111–1114; [18] Figueroa, W.G., Raszkowski, R. & Weiss, W.

Table 1.5 *(contd)*

(1973) Lung cancer in chloromethyl methyl ether workers. *New Engl. J. Med.*, **288**, 1096–1097; [19] Thiess, A.M., Hey, W. & Zeller, H. (1973) Zur Toxikologie von Dichlordimethyläther – Verdacht auf kanzerogene Wirkung auch beim Menschen. *Zbl. Arbeitsmed.*, **23**, 97–102; [20] Acheson, E.D., Cowdell, R.H. & Jolles, B. (1970) Nasal cancer in the Northamptonshire boot and shoe industry. *Br. med. J.*, **i**, 385–393; [21] Stott, H., Fox, W., Girling, D.J., Stephens, R.J. & Galton, D.A.G. (1977) Acute leukaemia after busulfan. *Br. med. J.*, **ii**, 1513–1517; [22] Reimer, R.R., Hoover, R., Fraumeni, J.F., Jr & Young, R.C. (1977) Acute leukemia after alkylating-agent therapy of ovarian cancer. *New Engl. J. Med.*, **297**, 177–181; [23] Thiede, T., Chievitz, E. & Christensen, B.C. (1964) Chlornaphazine as a bladder carcinogen. *Acta med. scand.*, **175**, 721–725; [24] Bidstrup, P.L. & Case, R.A.M. (1956) Carcinoma of the lung in workmen in the bichromates-producing industry in Great Britain. *Br. J. ind. Med.*, **13**, 260–264; [25] Smith, C.D., Prentice, R., Thompson, D.J. & Herrmann, W.L. (1975) Association of exogeneous estrogen and endometrial carcinoma. *New Engl. J. Med.*, **293**, 1164–1167; [26] Ziel, H.K. & Finkle, W.D. (1975) Increased risk of endometrial carcinoma among users of conjugated estrogens. *New Engl. J. Med.*, **293**, 1167–1170; [27] Mehnert, W.H., Haas, J.F., Kaldor, J., Day, N.E., Kittelmann, B., Stanaczek, W. & Möhner, M. (1986) *A case-control study of leukaemia as a second primary malignency following ovarian and breast neoplasms.* In: Schmähl, D. & Kaldor, J.M., eds, *Carcinogenicity of Alkylating Cytostatic Drugs (IARC Scientific Publications No.* **78**), Lyon, International Agency for Research on Cancer, pp. 203–221; [28] Herbst, A.L., Ulfelder, H. & Poskanzer, D.C. (1971) Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *New Engl. J. Med.*, **284**, 878–881; [29] Rang, E.H. & Acheson, E.D. (1981) Cancer in furniture workers. *Int. J. Epidemiol.*, **10**, 253–261; [30] Dublin, L. & Spiegelmann, M. (1947) The longevity and mortality of American physicians 1938–42. *J. Am. med. Assoc.*, **134**, 1211–1220; [31] Court Brown, W.M. & Abbatt, J.D. (1955) The incidence of leukaemia in ankylosing spondylitis treated with X-rays. *Lancet*, **i**, 1283–1287; [32] Lange, R.D., Moloney, W.C. & Yamawaki, T. (1954) Leukaemia in atomic bomb survivors. *Blood*, **9**, 574–580; [33] Hueper, W.C. (1966) Occupational and environmental cancers of the respiratory system. *Recent Results Cancer Res.*, **3**, 105–107; [34] Law, I.P. & Blom, J. (1977) Second malignancies in patients with multiple myeloma. *Oncology*, **34**, 20–24; [35] Stern, R.S., Zierler, S. & Parrish, J.A. (1980) Skin carcinoma in patients with psoriasis treated with topical tar and artificial ultraviolet radiation. *Lancet*, **i**, 732–735; [36] Wada, S., Nishimoto, Y., Miyanishi, M., Kambe, S. & Miller, R.W. (1968) Mustard gas as a cause of respiratory neoplasia in man. *Lancet*, **i**, 1161–1163; [37] Doll, R. (1958) Cancer of the lung and nose in nickel workers. *Br. J. ind. Med.*, **51**, 217–223; [38] Doll, R., Morgan, L.G. & Speizer, F.E. (1970) Cancers of the lung and nasal sinuses in nickel workers. *Br. J. Cancer*, **24**, 623–632; [39] Damon, A. (1960) Host factors in cancer of the breast and uterine cervix and corpus. *J. natl Cancer Inst.*, **24**, 483–516; [40] Friedman, G.D., Kannel, W.B. & Dawber, T.R. (1966) The epidemiology of gallbladder disease: observations in the Framingham study. *J. chron. Dis.*, **19**, 273–292; [41] Wynder, E.L., Cornfield, J., Schroff, P.D. *et al.* (1954) A study of environmental factors in carcinoma of the cervix. *Am. J. Obstet Gynecol.*, **68**, 1016–1047; [42] Pedersen-Bjergaard, J., Nissen, N.I., Sørensen, H.M., Hou-Jensen, K., Larsen, M.S., Ernst, P., Ersbøl, J., Knudtzon, S. & Rose, C. (1980) Acute non-lymphocytic leukemia in patients with ovarian carcinoma following long-term treatment with treosulphan (=dihydroxybusulfan). *Cancer*, **45**, 19–29; [43] Scotto, J., Fears, T.R. & Fraumeni, J.F., Jr (1982) *Solar radiation.* In: Schottenfeld, D. & Fraumeni, J.F., Jr, eds, *Cancer Epidemiology and Prevention*, Philadelphia, W.B. Saunders Co., pp. 255–276; [44] Creech, J.L., Jr & Johnson, M.N. (1974) Angiosarcoma of the liver in the manufacture of polyvinyl chloride. *J. occup. Med.*, **16**, 150–151; [45] Waxweiler, R.J., Stringer, W., Wagoner, J.K., Jones, J., Falk, H. & Carter, C. (1976) Neoplastic risk among workers exposed to vinyl chloride. *Ann. N.Y. Acad. Sci.*, **271**, 40–48; [46] MacMahon, B., Cole, P., Lin, T.M., Lowe, C.R., Mirra, A.P., Ravniher, B., Salber, E.J., Valaoras, V.G. & Yuasa, S. (1970) Age at first birth and cancer of the breast. A summary of an international study. *Bull. World Health Organ.*, **43**, 209–221; [47] Joly, D.J., Lilienfeld, A.M., Diamond, E.L. & Bross, I.D.J. (1974) An epidemiological study of the relationship of reproductive experience to cancer of the ovary. *Am. J. Epidemiol.*, **99**, 190–209; [48] Balarmaric, J. (1973) Intrahepatic bile duct carcinoma and *C. sinensis* inflection in Hong Kong. *Cancer*, **31**, 468–473; [49] Ferguson, A.R. (1911) Associated bilharziosis and primary malignant disease of the urinary bladder with observations in a series of forty cases. *J. Pathol. Bacteriol.*, **16**, 76–94; [50] de-Thé, G.B., Geser., A., Day, N.E., Tukei, P.M., Williams, E.H., Beri, D.P. Smith, P.G., Dean, A.G., Bornkamm, G.W., Feorino, P. & Henle, W. (1978) Epidemiological evidence for causal relationship between Epstein-Barr virus and Burkitt's lymphoma from Ugandan prospective study. *Nature*, **274**, 756–761

[b] The evidence comes from a wide variety of sources, and no single study can be regarded as definitive. The reference is to a review

compared to case-control studies are described as follows:

(a) A wider picture is obtained of the health hazards associated with a given exposure. This point was stressed when discussing the early studies on the effects of cigarette smoking (see Tables 1.2a and 1.2b). The link with the disease under prime suspicion, lung cancer, was established by retrospective case-control studies, but identification of the full range of diseases for which smoking increases the risk came from the prospective studies. Perhaps the most comprehensive longitudinal study of an exposed population with cancer as a major endpoint of interest is the follow-up of the survivors of the atomic bomb explosions in Japan (see Appendix IB for a description of the study design). An excess of leukaemia had been identified before the main programme started, but it was expressly stated when the atomic bomb survivor studies were launched that the overall aim was to study all the long-term health effects of

Table 1.6 Agents for which quantitative information is available on risk and exposure

Agent	Site	Type of study providing the principal information	
		Quantitative information on level of exposure	Quantitative information on temporal development
Cigarette smoking	Lung	Case-control and cohort	Cohort
	Other sites	Case-control	Little available
Alcohol	Oesophagus	Case-control	Little available
Asbestos	Lung	Cohort	Cohort
	Mesothelioma	Cohort	Cohort
Ionizing radiation	Most sites	Cohort	Cohort
Conjugated oestrogens	Endometrium	Case-control	Case-control

ionizing radiation, and in particular to determine if there was any evidence for a general acceleration of ageing. The enormous value of the study in providing the most precise estimates of the radiogenic cancer risk that are currently available for the majority of sites (Committee on the Biological Effects of Ionizing Radiation, 1980) has tended to obscure the major, if negative, findings that no detectable increase in mortality rates for nonmalignant diseases has occurred, nor is there evidence for an acceleration of the ageing process. The full picture of the long-term health effects of a given exposure can be provided only by the cohort approach.

(b) Recall and selection bias can usually be eliminated. This was perhaps the principal reason for launching the major prospective studies of cigarette smoking. Recall bias, a bugbear of case-control studies, should not occur in cohort studies. It is sufficient, of course, that recall bias could have occurred, rather than that it demonstrably did, for the results of a study to be questioned. An illustrative example of the doubts that may surround information that is obtained retrospectively is the early report of an excess of cancer among children irradiated *in utero* for diagnostic purposes (Stewart *et al.*, 1958). This study was based on interviews of mothers of cases and of controls about their pregnancy history after diagnosis of cancer in the case child. It was initially discounted because of the possible recall bias on the part of the mother – a criticism almost impossible to refute from within the study. A cohort study involving all births between the years 1947 and 1954 in the principal Massachusetts maternity hospitals, undertaken to test the validity of the association, gave quantitatively similar results (MacMahon, 1962). Since the cohort approach of this latter study avoids recall bias, any bias in the results must derive from the selection of women who undergo diagnostic X-ray during pregnancy and not from problems of recall. More detailed examination of the association in terms of year of irradiation (Bithell & Stewart, 1975), and more recent studies of twins (Boice *et al.*, 1985), suggest that the association is real.

Bias is not the only way in which differences in recall between cases and controls can

Table 1.7 Effect of different precision of response between cases and controls, for a polytomous exposure variable

	Levels of exposure					
	0	1	2	3	4	Total
'True' distribution (in cases and controls)	0.05	0.3	0.3	0.3	0.05	1.00
Observed distribution[a] in cases	0.075	0.275	0.3	0.275	0.075	1.00
Observed distribution[b] in controls	0.10	0.25	0.3	0.25	0.10	1.00
Apparent odds ratio	1.0	1.47	1.33	1.47	1.0	
			1.37			

[a] Obtained by spreading 10% of the true distribution on each side of the correct data point

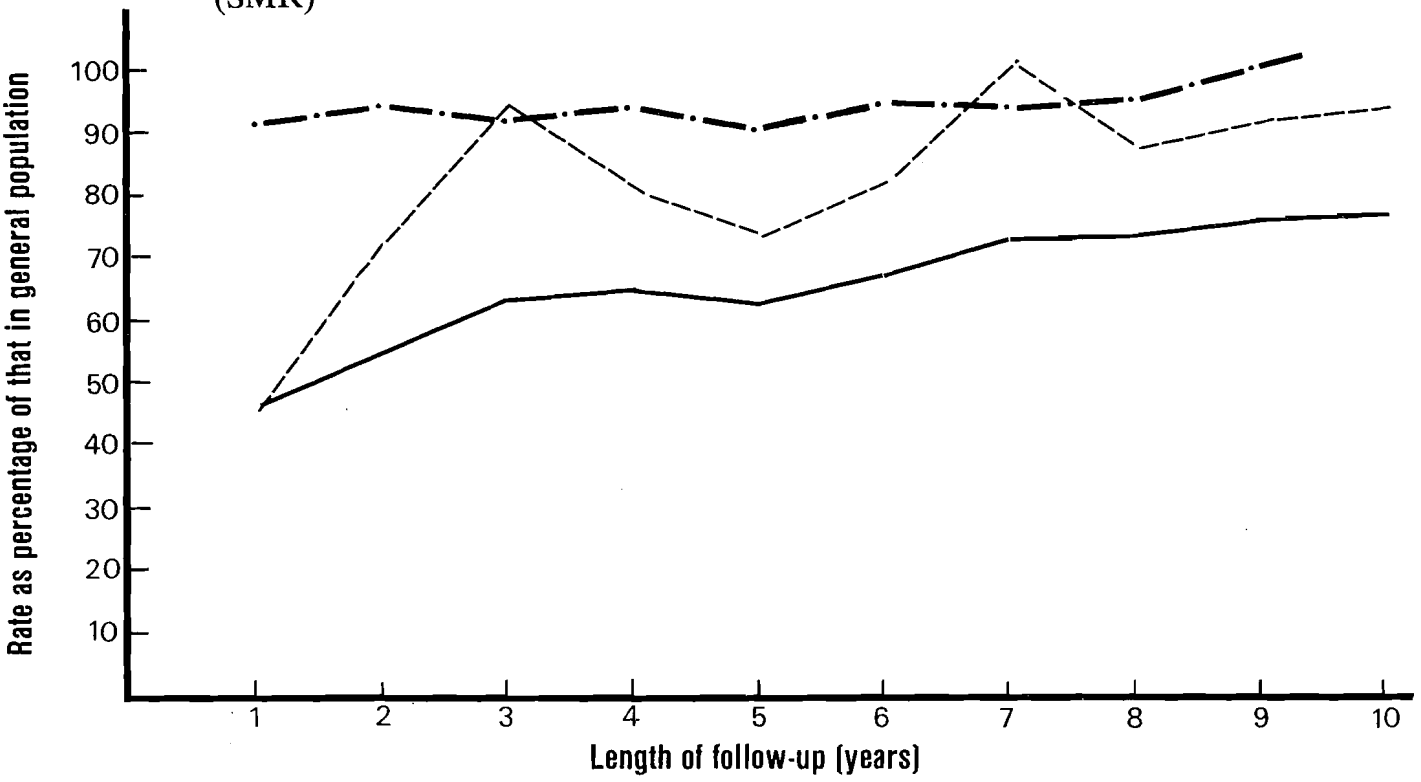
[b] Obtained by spreading 20% of the true distribution on each side of the correct data point

distort risk estimates derived from case-control studies, although it may be the major one. Distortion can also arise if the precision of the recall is different between cases and controls, again not a problem that should arise in a properly conducted cohort study. Consider a situation such as that illustrated in Table 1.7, in which the true distribution of a polytomous response is the same in cases and controls. Among both cases and controls, there is some unbiased random error in response, with a greater standard error among the controls. The effect is to generate a spurious risk differential. The picture of apparent low risk among a nonexposed group of low frequency, and lack of dose-response among the exposed is similar to that seen, for example, in several studies associating coffee drinking with bladder cancer (Hartge *et al.*, 1983).

Selection bias in case-control studies is often almost impossible to evaluate. If population controls are used, a large proportion of those originally selected may refuse to participate; if hospital controls are used, the choice of which disease categories to include is difficult to resolve, particularly for complex exposures like diet. It is rare for the case series to approach 100% of those arising in a defined population; if the series is eventually matched to a cancer registry, the proportion of eligible cases actually included seldom approaches the coverage of 90% or more considered acceptable for cohort studies (see Table 1.10). For cohort studies with good follow-up, problems of selection bias should not arise.

Provided that the follow-up mechanisms do not favour particular exposure groups, which can be checked by examining the data, then comparison of the disease experience among different subgroups of the study cohort should be unbiased. The cohort itself, however, will usually be a selected subgroup of the general population, and the disease experiences of the cohort and that of the general population may well not be comparable. The best known example of this lack of comparability is the so-called 'healthy worker effect'. The employed population is generally healthier than the nonemployed population of the same age, and their death rates for many causes of

Fig. 1.2 Evolution of the 'healthy worker effect' following entry into the study: Swedish building workers. The permission of Drs A. Englund and G. Engholm to reproduce this figure is gratefully acknowledged. ·—· Cancer incidence (SIR); – – – cancer mortality (SMR); ——— total mortality (SMR)



death are lower than the corresponding rates in the general population (Fox & Goldblatt, 1982). Cancer death rates appear to suffer less from the healthy worker effect than rates from most other causes, and cancer incidence rates probably are less affected than cancer death rates (see Figure 1.2).

The cohort may also be a selected subgroup of those exposed to the agent of interest. In both the Welsh nickel workers and Montana smelter workers studies considered throughout this monograph (see Appendices ID and IE), exposure was markedly higher before 1925 than in later periods. To qualify as a cohort member, however, individuals had to have been employed at the respective plant at some date in the 1930s or later. It may well be that employees not included, – for example, itinerant workers given the dirtiest jobs or those retiring early for reasons of health – differ both in their exposure and in their response from those included in the study. The cohort would then represent a biased selection of all those employed. This bias does not of course affect the inferences that are applied to the study cohort, but does affect the generalizability of the conclusions to all those working at the plant.

(c) Effects are more efficiently studied of exposures that are both rare in the general population and responsible for only a small proportion of any specific cancer. For many of the agents listed in Table 1.5, the cohort approach was taken to study the possible carcinogenic hazard because the exposure was rare. Only a very small proportion of

cases in the general population would have been associated with it, and a population-based case-control study would contain too few exposed cases to be informative. Cohort studies are sometimes claimed to be the approach of choice whenever an exposure is rare, but if a rare exposure is responsible for a substantial proportion of the cases of some cancer, presumably itself rare, then a case-control approach may well be more informative. Thus, the evidence relating adenocarcinomas of the vagina in young women to prenatal exposure to diethylstilboestrol, in general a rare exposure, comes essentially from case-control studies. In at least one prospective study of individuals exposed transplacentally to diethylstilboestrol no case of the disease was observed because, although the relative risk is very high, the baseline incidence from which the relative risk is calculated is extremely low. The associations of asbestos with mesothelioma and of vinyl chloride with angiosarcoma of the liver are also clearly demonstrated by a case-control approach, although quantitative aspects of the relationship may not be well estimated.

(d) Pre-disease information on biological parameters is available. Many physiological or biochemical measures, for example of nutritional status, will be modified by the disease itself, and observations taken on cases would be of doubtful value for etiological studies. As an illustrative example, in the early and mid-1970s, a series of case-control studies of hepatocellular carcinoma was performed investigating the differences in the prevalence of the carrier state of the hepatitis B virus surface antigen between the two groups. A much greater prevalence of the carrier state was seen among the cases in studies from a number of different countries (Szmuness, 1978). The etiological significance of this difference, however, was not accepted until it had been shown in a prospective study from Taiwan (Beasley *et al.*, 1981) that the same differences in rates between carriers and noncarriers were observed even when the carrier state was ascertained years before the onset of disease. The findings of several prospective studies have now been published, confirming the Taiwan study, and the association is regarded as definitively established.

A second example is given by the present great interest in the role of micronutrients in cancer etiology. Since retrospective recall of diet gives only a weak indication of intake of specific micronutrients, attention has been focused on physiological measures of levels of vitamins and trace elements. For these measures to reflect etiology, rather than simply the presence of disease, the requisite biological samples have to be taken some time before onset of disease, thus necessitating a prospective cohort approach. Existing banks of biological materials can be of great value in this context, forming potentially the basis for historical cohort studies. As frequently happens with material being used for a purpose not envisaged when it was collected, the condition of the collection and storage may not have been optimal, may even have been inadequate, for the parameters of current interest. Care must also be taken to ensure that samples from cases are not manipulated, for example thawed and refrozen, more frequently than the samples taken from the rest of the cohort.

(e) Information obtained retrospectively may be essentially too inaccurate to be of use. Exposure to aflatoxin, say, might be estimated by dietary recall combined with tables of aflatoxin levels in common foodstuffs, but one would not expect such

estimates to be accurate. The only way to obtain accurate measures of exposure is by direct observation of aflatoxin intake, either by assaying diet or measuring aflatoxin levels in body fluids. In another setting, quantitative estimates of exposure to benzene might be attempted by means of job histories linked to occupation-specific environmental measures, but better estimates of exposure are again clearly obtained by direct observation. The difficulty of interpreting past exposure measurements is well illustrated by the correspondence following the study of Infante *et al.* (1977) on benzene and leukaemia. The prospective study of gas workers in the United Kingdom, reported in 1965 and 1972 (Doll *et al.*, 1965, 1972), was initiated in 1953 because job histories could not be obtained retrospectively with acceptable accuracy. Assessment of exposure by biological monitoring has received increasing attention in recent years. Levels in the blood or in the urine of the compound itself or its metabolites can give measures of individual exposures, and this approach holds great promise for future prospective studies (Vainio, 1985).

(f) Repeated measurements can potentially be obtained. With an extended period of follow-up, serial measurements of the exposure may be possible. One consequence is that variables that have a time component are more accurately determined. A second consequence is that one can study the effect of changing levels of exposure. In breast cancer, for example, changes in weight may be of as much importance as weight just prior to diagnosis. Studying the effect of changing exposure levels provides one directly with estimates of the effect of intervention, although without the element of randomization. A third consequence is that misclassification rates or errors of measurement of exposures and of confounding variables can be assessed in unbiased fashion for the general population and for individuals destined to develop the disease of interest. These estimates can be used to infer the real rather than the apparent effect of exposure (Clayton & Kaldor, 1985). The use of prospective studies to generate repeat measurements of the variables of interest, and to use these repeat measurements to obtain estimates of the real disease-exposure relationship, is an area worth further elaboration.

(g) For some purposes, one requires not relative but absolute measures of risk. Cohort studies provide direct estimates of incidence rates, as opposed to the ratios of rates estimable from case-control studies. Both for public health decisions and for the study of the mechanisms of carcinogenesis, incidence rates, giving as they do absolute measures of risk, may sometimes be preferable.

1.3 Limitations of cohort studies

In spite of their advantages, a history of cancer epidemiology indicates that cohort studies have not been the major avenue of attack. Case-control studies have predominated. The limitations of cohort studies are summarized below.

(a) Prospective cohort studies imply a commitment over many years, and both individuals and granting agencies are loath to embark on a project that will not yield its main results for a decade or more. Furthermore, collecting accurate information on more than a short set of variables from the large number of individuals required for a

cohort study may be very expensive. The use of case-control comparisons within a cohort (see §1.4*i* and Chapter 7) may reduce the workload involved in processing the data, but the costs of collecting the data still have to be taken into account.

(b) Historical cohort studies do not suffer from this extended commitment into the future, but they can obviously be performed only if a relevant cohort can be identified. For many exposures, the existence of such a cohort with accurate records of exposure dating back ten or more years cannot be guaranteed. Furthermore, even if a cohort which seems to approximate to the requirements of the study can be identified historically, information on other variables which may play an important confounding role is likely to be lacking. Thus, one can study diabetics as a group likely to consume greater than average quantities of artificial sweeteners, but the incidence of bladder cancer in the group is difficult to evaluate in the absence of information on smoking – less common among diabetics. The difficulty of evaluating the significance of moderate excesses of lung cancer among industrial cohorts for whom smoking histories are not recorded is well known.

(c) Most cancers are rare diseases. Even taking one of the most frequent cancers – breast cancer among women over 50 years of age from western countries – no more than one or two cancer cases could be expected per 1000 women per year. For most other cancers, the expected numbers would be considerably less. In fact, for rare cancers, cohort studies are unlikely to be of much value, unless the relative risk associated with the exposure under study is very large. Questions of power are considered in some detail in Chapter 7, but as a noteworthy illustration of the point, in the prospective study of British doctors, the association of smoking with bladder cancer, notwithstanding an observed increase in risk for current smokers of twofold, was not significant. Bladder cancer, although not one of the most common cancers in the United Kingdom, is also not one of the most rare, but nevertheless the number of cases was insufficient to demonstrate the effect convincingly. Given the relative ease with which a series of several hundred cases of cancer at many particular sites can be assembled, it is clear that on most occasions a comparison of the effort or cost per case included in the study will favour the case-control rather than the cohort approach. The justification for a cohort study has to be based, usually, on the superiority of the information it can yield.

(d) It is difficult to obtain estimates of attributable risk. On many occasions, there is interest not only in the degree of risk associated with a certain exposure, but also the importance of that risk in the general population. For a given cancer site, this can be expressed as the population attributable risk, a quantity which population-based case-control studies can provide in straightforward fashion. Many cohort studies, however, are based purposely on groups with a much higher prevalence of the relevant exposure than the general population and so cannot give estimates of the population attributable risk; when the results are extrapolated to the general population, bizarre conclusions can be drawn (as for example the unpublished but widely quoted report from the US National Institutes of Health on the proportion of cancers due to occupational exposures). Even cohorts such as the British doctors or US veterans differ sufficiently from the general population, in terms of economic level for example, to make extrapolation in terms of attributable risk hazardous.

The preceding discussion of the merits and drawbacks of a cohort study relative to a retrospective case-control study suggests that the two approaches have complementary attractions. The cohort approach provides a well-defined population from which cases can be identified in an unbiased manner, and for which some contemporary information on the exposure of interest will be available throughout the time period of interest. Time variables for this exposure may also be well recorded. The case-control approach concentrates effort on the informative individuals, that is the cases and a suitable set of controls, on whom extensive information on confounding variables may be obtainable. In an increasing number of situations, study designs are being used that incorporate the case-control approach within a cohort study, to take advantage of the merits of both approaches. These designs are discussed further in §1.4*i* and in detail in Chapters 5 and 7.

1.4 Implementation

The major concern in assembling a cohort for study is that the size and level of exposure of the cohort be sufficient to yield meaningful results. The two questions that arise are, first, is the study worth doing and, second, how should it be implemented. Chapter 7 considers the question of statistical power, a major element in assessing the potential informativeness of a study. In this section, we treat a range of issues which arise in the implementation of a study.

The design and execution of a cohort study will depend on the individual circumstances of the study, and on its aim. It is instructive to examine in detail the methods used in a number of studies, and for this purpose we give an extended account, in Appendices IA–IF, of the implementation of different types of cohort study – six in all. The studies vary widely in many respects. The study of atomic bonb survivors has been the largest single programme of research in chronic disease epidemiology. The Life-Span Study, on which the majority of the mortality results are based, forms both an infrastructure from which a broad range of activities has evolved and the main source of information on the cancer risks associated with radiation available at the present time (Committee on the Biological Effects of Ionizing Radiation, 1980). An enormous effort has gone into defining the cohort and estimating the radiation dose received by each member of the cohort, and to ensuring the completeness of follow-up and the validity of the death certificate information. All possibly fatal consequences of radiation exposure were included in the scope of the study.

By contrast, the hepatitis B prospective study was targeted closely on the association between the hepatitis B surface antigen carrier state and risk of primary liver cancer. A particular effort was made to confirm primary hepatocellular carcinoma as the cause of death, but otherwise the formation of the cohort and the follow-up procedures depended mainly on existing facilities.

Even though the scope and purpose of different studies may vary widely, however, there are a number of issues in the design and execution that require attention, irrespective of whether the study is prospective or historical. These questions include

the following:

(*a*) Who should be included?
(*b*) What dates should be taken for each individual as the date of their entry into the study, and their date of exit from the study?
(*c*) What follow-up mechanisms or systems are to be used, and to what extent will demands of confidentiality impinge on the completeness of follow-up data?
(*d*) What endpoints will be used for assessing disease occurrence, and how are disease categories to be coded?
(*e*) What information should be obtained on exposure, and how often during the period of follow-up can it, and will it, be assessed?
(*f*) What information is available on other exposures, which are either of intrinsic interest or of importance as potential confounding factors?
(*g*) What comparisons will be made to assess the effect of exposure?
(*h*) What power will the study have to detect the levels of excess risk that might realistically be expected?
(*i*) Can a case-control approach be introduced into any component of the study, to improve feasibility or reduce costs, without reducing the information content?

(*a*) *and* (*b*) *Who should be included, and how are dates of entry and exit defined?*

The main requirement is that inclusion rules be clear and unambiguous. For each individual, the date on which observation begins must be well defined; after that date the individual contributes person-years of observation and is at risk to contribute events of interest. Table 1.8 summarizes the definition of the study cohort for the six studies described in detail in Appendices IA–IF. For each study, it is clear who is a cohort member, and from what date cohort membership starts. If should be stressed that an individual does not enter the cohort, and contribute person-years at risk, until all entry criteria have been satisfied. Thus, in the South Wales nickel workers study, an individual is included and starts to contribute person-years of exposure at the date on which his second appearance on a pay sheet occurs, not on the date of his first appearance on a pay sheet.

The date of exit from the study is the last date on which an individual could contribute person-years at risk. In every study, a date has to be specified as the end of the follow-up period for the current analysis. The vital status on that date should be ascertained for all cohort members and explicitly tabulated when reporting the study. In the Montana smelter workers study, the follow-up was originally to 31 December 1963, and later extended to 30 September 1977. Follow-up status in the original and the extended study is displayed in Tables 1.9a and 1.9b. For those whose vital status was known at the given date, the date of exit from the study is that date or the date of death – whichever is the earlier. Those whose vital status is not known at the end of follow-up will have been lost to follow-up, and the correct procedure is to terminate their follow-up on the last date their vital status is known.

It is important to note that date of entry into the study and date of first exposure are not necessarily the same, and will often be different. In the South Wales nickel workers

Table 1.8 Cohort definition in the studies described in Appendices IA–IF

Study	Source records used to define cohort	Inclusion criteria	Date of entry into cohort (i.e., into follow-up for calculation of person-years)
British doctors	UK medical register	Satisfactory reply to a mailed questionnaire, posted on 31 October 1951	1 November 1951
Atomic bomb survivors	National census data, 1 October 1950	Japanese citizens, present in city at time of bomb, resident in city at census date, place of family registration in or near city (later relaxed)[a]. All individuals within 2500 m of bomb hypocentre ATB, plus a sample of those more distant than 2500 m	1 October 1950
Taiwan hepatitis study	Taiwan government employees	Presented themselves at Government Employees' Clinic Centre for a routine free health examination, and agreed to provide an extra sample of blood	Date of visit to Clinic
South Wales nickel refinery	Company employment records (weekly pay sheets) for first week of April, 1934, 1939, 1944, 1949	Included[b] on at least two of the four pay sheets used	Date of the second pay sheet on which individual appears
Montana smelter workers	Company employment records	Employed for at least one year before 31 December 1956	At end of 1 year's employment or 1 January 1938 if employed for more than 1 year before that date
North American asbestos workers	Union membership records, 1966	All union members in 1966, alive on 1 January 1967, for asbestos study; asbestos and smoking study limited to 11 656 individuals who completed and returned a questionnaire on smoking habits	1 January 1967

[a] When the original cohort was assembled, a 'reserve' group was included who satisfied all criteria except that their place of family registration was distant from either city. This 'reserve' group was added to the rest of the cohort during the follow-up period (Beebe *et al.*, 1977)

[b] This criterion refers to the original cohort described in the first paper (Doll *et al.*, 1970). The cohort was later extended

Table 1.9a Status of study group, 31 December 1963[a]

Known to be living		5397
Employed by smelters	1471	
Pensioned	389	
Other[b]	3537	
Known to be deceased		1877
Vital status not known		773
Total study group		8047

[a] From Lee and Fraumeni (1969)
[b] Includes persons receiving benefits or making claims to Bureau of Old Age and Survivors Insurance after December 1964

Table 1.9b Follow-up status of study group[a]

Follow-up status, 30 September 1977	Total study group, 1977	Follow-up status, 31 December 1963		
		Known to be living	Known to be deceased	Vital status not known
Known to be living	3707[b]	3342	0	365[c]
Known to be deceased	3522	1534	1877	111
Vital status not known	816	520	0	296
Total, 1963	8045[d]	5396	1877	772

[a] From Lee-Feldstein (1983)
[b] Includes 442 men still employed at the smelter on 30 September 1977
[c] Approximately half of the men reported lost to follow-up by Lee and Fraumeni (1969) were found to be alive on 30 September 1977
[d] Two persons in the original study group of 8047 were women; they have been deleted from the present study

study, at least five years had to elapse after first exposure before entry into the period of observation. For many in the cohort, employment began ten, 20 or even 30 years before follow-up began. Among the atomic bomb survivors, five years elapsed between exposure and the census that defined the cohort. In some studies, it is possible to ensure that follow-up begins as soon as exposure occurs, but this is not usually so. A distinction has been made, mainly of interest in the occupational setting, between so-called 'prevalence cohorts', consisting of all those employed on a particular date, and 'incidence cohorts', consisting of all those first employed between two dates. The possible problems of interpretation when using prevalence cohorts have been mentioned earlier.

(c) *Follow-up mechanisms*

Follow-up over time of the individuals enrolled in a cohort study is the essential feature of the study. The success with which the follow-up is achieved is probably the basic measure of the quality of the study. If a substantial proportion of the cohort is lost to follow-up, the validity of the study's conclusions is seriously called into question.

Table 1.10 Completeness of follow-up in a number of cohort studies

Study	Date of end of follow-up	Proportion with unknown vital status at that date
British doctors	Males: 1 November 1971 Females: 1 November 1973	103 (0.3%) from 34 440 61 (1.0%) from 6194 (Authors estimate that at most 'a dozen' out of some 10 000 deaths would have been missed)
Atomic bomb survivors	31 December 1978	The follow-up depends on the national family registration system. Only 9 (0.7%) of 1300 known deaths were not recorded by the system in a pilot investigation (Beebe *et al.*, 1962). Losses to the cohort from emigration were estimated to be less than 100 (0.1%) by the end of 1974 (Beebe *et al.*, 1977)
Taiwan hepatitis study		Passive surveillance: 74 (0.3%) from 22 707 Active surveillance: 74 (1.1%) from 6908
South Wales nickel refinery workers	31 December 1971	37 (3.8%) from 967
Montana smelter workers	30 September 1977	816 (10.1%) from 8045
North American asbestos workers	31 December 1976	100% follow-up, according to the investigators (Hammond *et al.*, 1979)

The loss to follow-up reported in the studies described in appendices IA–IF is shown in Table 1.10, and indicates the target to be achieved. It is worth repeating the point made in an earlier section that case-control studies, insofar as they can be interpreted in a cohort context, would rarely achieve comparable coverage. In situations where new or untried follow-up mechanisms are to be used, a pilot study of their efficacy is recommended, as performed in the atomic bomb survivors study (see Table 1.10).

The purposes of the follow-up process are threefold:

(i) to determine which cohort members are currently under observation, by recording deaths and losses due to migration, i.e., to determine the denominator information;

(ii) to determine the disease events that are the defined endpoints of the study, i.e., to determine the numerator information; and

(iii) to obtain further information on the cohort members.

(i) *Determination of denominator information*

The mechanisms to be used for follow-up vary from country to country, depending on the national systems of population registration and on local laws on confidentiality. If the cohort is defined in terms of an occupational or professional group, or membership of a health insurance plan, then these group records often provide an accurate mechanism for follow-up. In Scandinavian countries, each individual has a unique identifying number which is in common use, accurate records of death using this number extend back over decades, and population rosters exist listing all persons presently living in the country. In principle one can therefore ascertain from among

cohort members who has died in the country, who is currently alive in the country and thus, by default, who has left the country at some time in the past. This last group can be investigated further to determine the date of exit from observation. In England and Wales, linkage to death records can usually be achieved easily and with little error, but verification that the individual is currently alive and living in England and Wales is in general cumbersome. In the British doctors study, attempts were made to contact each individual at the end of the follow-up period to ascertain vital status. A range of mechanisms is available in other countries, which is of varying utility for cohort studies. In some countries, the contents of a death certificate are confidential, and it is illegal to know the cause of death.

Since many cohorts under study are defined in terms of membership of a particular group (professional body, occupational pension plan, union, health insurance plan, college alumnae), then, provided individuals remain members of the group, supplementary information from group records may be available, particularly on present vital status. For some cohorts, records kept by the group may be superior to the national system. The British doctors study, the Taiwan hepatitis study and the American insulators study used group records to ascertain vital status, and, as can be seen in Table 1.10, the losses to follow-up were small.

An issue that needs special attention if group membership is related to employment is the question of retirement. Loss to follow-up among the retired, if it achieves appreciable proportions, can vitiate a study, particularly since retirement can be caused by ill-health. Increased mortality soon after retirement is a common observation. Efforts made to trace those who had retired received special attention in the report on the Taiwan hepatitis study (see Appendix IC).

On occasion, follow-up may be active in the sense that the investigators attempt to see each cohort member on a regular basis. Such an approach can clearly be expensive, and it has tended to be used mainly when the cohort is already under some form of clinical care. Thus, in an early study of women irradiated for cancer of the cervix, on the routine post-treatment visits made by the cervical cancer patient to her treatment clinic, further investigations were carried out as part of the study – notably, haematological studies, since leukaemia was the main endpoint of interest at that time (Hutchison, 1968). In the Taiwan study, hepatitis B surface antigen carriers and a negative group of equal size were followed actively, with an annual examination. This active follow-up was undertaken partly to verify the completeness of the passive follow-up mechanism, and partly for sequential determination of hepatitis B virus status.

On occasion, acquisition of information on death and migration for every member of the cohort may involve greater expense than the study can meet. If the endpoints of interest can be ascertained on all cohort members, for example from a cancer registry, then a less complete approach could be considered to enumerate the relevant denominator. Two possibilities offer themselves. First, one can carry out the follow-up, for the purposes of person-year calculations, on a sample of the entire cohort. If the cohort is large, reasonably accurate denominator information may be obtained by sampling only a small fraction of the total. Second, one might use an actuarial approach and calculate what could be called 'expected' expected numbers, based on

population deaths and migration rates. Either approach might be considered in a pilot phase of the study, to determine whether a full study was of interest. Section 1.6 treats briefly the proportional mortality approach, where denominators are simply ignored.

(ii) *Identification of cancer cases among the cohort*

In many countries, use of cancer registries for ascertainment of cancer occurring in cohort members is limited by incomplete coverage on a national level. There is considerable risk that individuals will have moved to areas of the country not covered by cancer registration. In some countries, however, cancer registration is nationwide. Use of the cancer registry to ascertain the cancer cases occurring in the cohort then has a number of advantages over use of the usual alternative source of information, namely, death certificates.

First, recording of cancer in cancer registries is often more accurate than on death certificates. Greater care is taken to ensure that the registered diagnosis is correct and more information is given, particularly on the histology of the cancer.

Second, more cases should be observed, since many cancers do not lead to death. The cases will also be observed earlier, death occurring after diagnosis, often with a delay of several years.

Third, population rates will be available for a wider variety of cancers. Observed and expected numbers can be given, for example, for different histological types of lung cancer.

A fourth advantage, of relevance mainly in occupational studies, is that the healthy worker effect should be less for cancer incidence than for cancer mortality, and may in fact be almost negligible (as in Figure 1.1).

Notwithstanding the advantages of cancer registry material, a large proportion of cohort studies have no choice but to use death certification as the main source of information on cancer in the cohort. For all comparisons with national mortality rates, it is essential that the cause of death as given in the death certificate be used. Only in this way can unbiased comparisons be made.

(iii) *Further information on cohort members*

For analyses, however, in which subgroups of the cohort under study are compared, the diagnosis as given on the death certificate may benefit from refinement based on additional information. Centralized review of available histological or haematological material to ensure uniform classification of subtypes of disease, or simply to confirm death certificate diagnoses, can be a valuable exercise. The resulting reclassification might be expected to sharpen the analyses performed within the cohort. In the British doctors study, for example, confirmation was sought whenever lung cancer was given on the death certificate as the underlying or contributing cause of death.

In the study of North America insulators, it was considered likely that at least asbestos-related disease might have been misclassified. Further information, including histology sections and X-ray films, was obtained where possible on asbestos workers who had died, and a revised cause of death assigned whenever indicated. Table 1.11 shows the differences in the number of deaths due to different causes between the

Table 1.11 Number of deaths occurring from five through 35 years after onset of work in an amosite asbestos factory, 1941–1945. Cause of death coded in two different ways[a,b]

Underlying cause of death	DC	BE	BE–DC	Expected
All causes	1946	1946	—	1148.0
Cancer, all sites	845	912	+67	259.0
Lung	397	450	+53	81.7
Pleural mesothelioma	23	61	+38	—
Peritoneal mesothelioma	24	109	+85	—
Mesothelioma not specified above	54	0	−54	—
Larynx, buccal and pharynx	21	27	+6	7.5
Oesophagus	17	17	0	5.1
Kidney	15	16	+1	8.5
Colon-rectum	54	55	+1	8.5
Stomach	18	21	+3	30.5
Prostate	24	26	+2	12.5
Bladder	7	9	+2	6.7
Pancreas	46	21	−25	16.0
Other specified sites	110	83	−27	72.1
Primary site unknown	35	17	−18	
Noninfectious pulmonary diseases, total	177	204	+27	68.2
Asbestosis	76	160	+27	—
Cardiovascular disease	638	566	−72	660.1
Other and unspecified causes	286	264	−22	160.8
Subtotal, all causes except cardiovascular diseases	1308	1380	+72	487.9

[a] DC, cause of death according to death certificate information only; BE, cause of death according to best evidence available
[b] From Hammond *et al.* (1979)

death certification and the cause based on the best available evidence. The death certificates severely underestimate the number of deaths due to asbestosis, mesothelioma and, to a lesser extent, lung cancer. More accurate estimates of dose-response, and of changing risk with time, are clearly given by the diagnoses based on the best available evidence. It is of interest to note that pancreatic cancers were overdiagnosed on the death certificates, and the confusion was usually with mesothelioma. What appears as a greater than two-fold excess, highly significant statistically, disappears if the more accurate cause of death is used. (In this situation, one might legitimately compare the refined diagnosis with expectations based on death certificate diagnoses since the main source of error, mesothelioma, affects inappreciably the general population rates.)

In the Taiwan hepatitis study, clinical and pathology records were sought for all deaths that occurred in the cohort, and primary hepatocellular carcinoma accepted as the cause of death only if the evidence was unambiguous. This degree of confirmation was required since the aim of the study was precisely to define the risk of primary hepatocellular carcinoma among carriers and noncarriers.

(*d*) *Coding of disease*

For many cohort studies, the follow-up period may cover a period of several decades. During this time, the codes used both for death certification and for cancer registration have changed. Of the International Classification of Diseases (ICD), the 6th revision was introduced in 1950, the 7th in 1955, the 8th in 1968 and the 9th in 1978. The World Health Organization has asked member countries to code death certificates according to the current revision. Unfortunately, several disease categories are subdivided differently in different revisions, confusing the comparisons between time periods. The cancer section of the ICD has been disturbed in this way less than many other sections, and most investigations extending back to the 1950s or earlier have devised tables of equivalence between the different revisions. The equivalences given in *Cancer Incidence in Five Continents* (Waterhouse *et al.*, 1976) are reproduced in Appendix II.

(*e*) *Information on exposure, and how often it should be assessed*

The aims of the study should help to define the detail that is required for the information on exposure. In the studies described earlier by Case, exposure was described solely in qualitative terms, whether or not employment for more than six months had occurred in an occupation with recorded exposure to the compounds of interest (e.g., α- and β-naphthylamine, benzidine, aniline). The cohort was classified according to the compounds to which individuals had been exposed (see Table 1.4). No data were available to quantify the level of exposure. The data, however, were considered adequate for the purpose at hand, which was identification of the major bladder carcinogens in the chemical industry, and the results sufficed as a basis for legislation. Thirty years later, no more quantitative relationship between levels of exposure to the aromatic amines and bladder cancer risk in humans has been established.

In contrast, studies on the leukaemogenicity of benzene, although unequivocally demonstrating that benzene causes leukaemia in man (Infante *et al.*, 1977; IARC, 1982b), have been insufficient at present for societal purposes, i.e., for setting safety limits, because of uncertainty over the levels to which the cohorts had been exposed. Further studies on benzene would be of value only if quantitative exposure levels could be determined for each individual during his period of exposure.

However, although the degree of quantification possible under different circumstances varies, the more quantitative that one can make the relationship between exposure and risk, the more it will be of value, for three reasons. First, the credibility of the causality of the relationship will be enhanced; second, the greater will be the potential for meaningful public health action; and, third, the contribution to an understanding of carcinogenic mechansims will be increased. It should be stressed that quantifying exposure requires recording not only the level of exposure, but when it occurred, for how long and whether it stopped. These temporal aspects may well be more powerful determinants of risk (as discussed in Chapter 6) and are often better recorded than is the exposure level.

To obtain quantitative relationships between exposure and excess risk, information

Table 1.12 Information on exposures in the six studies described in Appendices 1A–1F

British doctors	Smoking history from questionnaires. Amount currently smoked and age at which smoking started and stopped (for ex-smokers). Questionnaires sent on four separate occasions
Atomic bomb survivors	Detailed information on individual dose for each member of the cohort
Taiwan hepatitis study	One determination of hepatitis B virus carrier status on entry to study
South Wales nickel refinery workers	Specific job held during each year of employment. Job posts categorized retrospectively by analysis of associated risk
Montana smelter workers	Time and place of employment for each job held within smelter. Environmental measures taken in the smelter enabled each work area to be categorized on a 1 to 10 scale for arsenic trioxide levels.
North American asbestos workers	No information on level of asbestos exposure. Start and length of employment. Smoking history obtained in 1966

on exposure at the individual level is of the greatest value. Although mean levels for the entire cohort are not valueless, since they do give some impression of what dose has produced a given excess risk, they do not reflect the fact that the study is individual-based, and cannot yield estimates of dose-response.

The extent and detail of the information on exposure should reflect the relationship between exposure and excess risk that the investigator might expect. Table 1.12 summarizes the exposure data collected in the six studies described in Appendices IA–IF. Quantitative models of carcinogenesis, considered in Chapter 6 in some detail, suggest some of the required information. First, in relation to time of exposure, for each individual one should know the dates at which exposure started and stopped and the subject's age when exposure started. It is not unrealistic to expect such information to be available if it is based on employment or medical records, or is derived from questionnaires. If the exposure information comes from biological markers, however, exposure status will be determined only for the time points when the relevant samples are taken. The design of the study will then be critically influenced by the hypotheses under test. For an example in which this issue is discussed in detail – a prospective study relating Burkitt's lymphoma to infection by the Epstein-Barr virus – see Geser and de-Thé (1972).

Second, in relation to level of exposure, quantitative information is rarely available throughout the period of exposure. Exceptions might be workers exposed to radioactivity, who are continuously monitored, or patients given chemotherapy, for whom details of treatment should be available. One has to decide which summary measures are most informative. For many exposures, average levels during the whole period of exposure may be sufficient. For asbestos and mesothelioma, however, the levels in the first few years of exposure are likely to be the most relevant and levels in the years immediately preceding disease onset almost irrelevant (Peto, J. *et al.*, 1981). In contrast, for cigarette smoking and lung cancer, Doll and Hill (1950) found from the results of their initial case-control study that among continuing smokers the amount most recently smoked was almost as informative as the full smoking history over many years.

Asking about current practices, or measuring current levels, has the advantage that it yields more accurate data than asking about former practices, although its value is restricted to prospective studies. It can also take advantage of new techniques for measuring metabolite levels in the urine, or binding to macromolecules in the blood, which offer the potential of measuring more relevant aspects of an individual's exposure. For many occupational or environmental exposures, however, levels have fallen steadily over the last three decades. For the determination of the present excess risk, the most relevant exposure levels may well be those in operation two or three decades ago, simply because at that time they were much higher. Unfortunately, if quantitative values are available for exposure levels 30 years ago, the methods used in their determination may not be comparable with those used today, or may not even be interpretable with more stringent modern criteria. Asbestos measurements, for example, taken in the 1940s are difficult to calibrate with modern measurements.

On many occasions the specific carcinogen may not have been identified, as for example in nickel refining, leather and wood working, or arsenic exposure in non-ferrous metal smelters. In these circumstances, dose cannot be defined in an absolute sense. One may, however, be able to assign degrees of exposure. Such a procedure was used when studying the Montana smelters exposed to arsenic, and in the most recent publications on the study of the South Wales nickel workers (Peto, J. *et al.*, 1984; Kaldor *et al.*, 1986).

In the Montana study, measurements of atmospheric arsenic trioxide were used to categorize working areas as providing heavy, medium or light arsenic exposure (Lee & Fraumeni, 1969). Each individual could then be categorized in terms of the jobs he had held within the smelter since the start of his empolyment there. It was noted that arsenic trioxide measurements were not made throughout the period of exposure, and that levels may have varied, but the authors considered that the relative exposure levels in different work areas would have remained fairly constant. In the nickel study, categorization of work areas was done *post hoc,* a high risk being associated with working in only a few of the refinery work places.

The degree of exposure defined categorically in this manner can then be used as a nonquantitative ordered variable in the analysis, affording the possibility of demonstrating a positive dose-response relationship. It may often happen that the categorization of level of exposure that gives the clearest trend with increasing risk will incorporate aspects of duration of exposure.

In prospective studies, the opportunity exists to repeat measurements of exposure, so that the degree of measurement error or of intrinsic intra-individual variation can be estimated. These estimates can be used to assess the real exposure-disease relationship (Clayton & Kaldor, 1985), but the planning of data collection to provide genuine estimates of intra-individual variation has at present received little attention in the context of prospective epidemiological studies.

(*f*) *Information on other exposures*

The main weakness of many historical cohort studies is the absence of information on potentially confounding variables. The lack will increase the uncertainty of the

interpretation placed on the results. Two possible approaches could be taken to improve the situation. First, for the relevant cases and a series of controls, one could mount an intensive effort to obtain the missing information. For a historical cohort study that has involved several decades of follow-up, many of the cases may be long since dead. One could not expect a high degree of accuracy from a surviving spouse or friend on, say, the smoking and drinking habits of someone who had died 15 or more years previously. Limiting this case-control accrual of information to cohort members alive within five years of the interview should improve the accuracy of the information, but may remove most of the cases. A second approach would be to take a sample of the surviving members of the cohort and obtain the necessary information from them. This information cannot, of course, be handled in the usual way for information on confounding variables, by stratification or incorporation in regression models, but it can be used to give an estimate of the degree of confounding associated with that variable. This estimate has to involve use of external information on the risk associated with the confounding variable. The confounding risk ratio would then be calculated as in Chapter 3 of Volume 1.

In prospective cohort studies, one would expect to obtain prospectively some information on potential major confounding variables.

For both historical and prospective cohort studies, it is important that the information obtained on confounding variables be reasonably accurate. Approximate information, such as one might feel appropriate for factors of secondary inportance, can be almost useless (Tzonou *et al.*, 1986). An example is given in Table 1.13, for a dichotomous exposure and a single dichotomous confounding variable, the latter being observed with error. The level of confounding that remains after the effect of the misclassified confounding variable has been taken into account is tabulated for a range of situations. Misclassification rates of 30%, not unknown in epidemiology, allow the removal of very little of the confounding effect; rates of 10%, which would often be considered relatively precise epidemiological measures, leave nearly half the confounding effect in operation. It is clear that one should attempt to obtain information prospectively on the entire cohort only if it is both economic and feasible to collect accurate information. Otherwise, the resources are probably better allocated to obtaining the information accurately on a case-control basis, concurrently if possible. An alternative approach, being developed, is to obtain repeat joint measures of the exposure and the confounder, and use the estimates of the joint error distribution to estimate the real relationship (Clayton & Kaldor, 1987).

(*g*) *The need for the construction of special comparison groups*

In most studies, the comparisons of interest that will be made are either among subgroups of the cohort or with the general population. On occasion, however, comparisons with an external group or among subgroups within the cohort will be insufficient. A separate control group will then have to be constructed. Such a situation is seen in the study of insulators, in which the emphasis is on the combined effect of smoking and asbestos exposure. To assess the effect of asbestos exposure among smokers and nonsmokers separately, one requires mortality rates among smokers and

Table 1.13 Bias in the estimation of the summary odds ratio if the confounding variable, *C*, is misclassified but the exposure variable, *E*, is not. The body of the table shows Δ_E, the ratio of the measured odds ratio to the true odds ratio (R_E)

P	p_1	p_2	$R_C = 2$				$R_C = 10$			
			$\delta = \gamma =$ 0.1	0.2	0.3	W	$\delta = \gamma =$ 0.1	0.2	0.3	W
0.1	0.6	0.4	1.05	1.09	1.12	1.14	1.15	1.26	1.34	1.39
	0.7	0.3	1.11	1.20	1.26	1.31	1.34	1.61	1.81	1.97
	0.8	0.2	1.20	1.34	1.43	1.50	1.62	2.16	2.58	2.93
	0.9	0.1	1.36	1.55	1.66	1.73	2.29	3.35	4.14	4.79
0.5	0.6	0.4	1.05	1.09	1.12	1.14	1.17	1.27	1.34	1.39
	0.7	0.3	1.12	1.20	1.26	1.31	1.40	1.67	1.84	1.97
	0.8	0.2	1.22	1.35	1.44	1.50	1.79	2.33	2.67	2.93
	0.9	0.1	1.41	1.58	1.67	1.73	2.74	3.76	4.36	4.79
0.9	0.6	0.4	1.05	1.09	1.12	1.14	1.18	1.28	1.35	1.39
	0.7	0.3	1.13	1.21	1.27	1.31	1.46	1.72	1.87	1.97
	0.8	0.2	1.23	1.37	1.45	1.50	2.00	2.49	2.75	2.93
	0.9	0.1	1.44	1.60	1.68	1.73	3.32	4.16	4.55	4.79

R_E = true odds ratio between exposure *E* and disease
P = (true) proportion of population exposed to *E*
R_C = true odds ratio between exposure *C* and disease
p_1 = proportion of those exposed to *E* who are also exposed to *C*
p_2 = proportion of those not exposed to *E* who are exposed to *C*
δ = proportion of those truly *C*+ classified as *C*−
γ = proportion of those truly *C*− classified as *C*+
W = confounding risk ratio (for estimation of R_E) (= estimate of odds ratio ignoring *C*/true odds ratio)

nonsmokers without asbestos exposure. Since the entire group of insulators was considered to have been exposed to asbestos, this requirement necessitated the construction of an ad-hoc comparison group. The procedure is described in Appendix IF, where it can be seen that considerable care was taken to match the comparison group on socioeconomic and other factors. It is interesting to note that a nonexposed control group was also assembled in the atomic bomb survivor studies, consisting of people present in the two cities at the 1950 census but not at the time of the bomb. Fears were expressed at the outset that this group might not be comparable in a number of respects, and it emerged that their mortality rates differed from those of the study cohort in ways unrelated to exposure. They were not included in most of the major analyses.

(*h*) *Power considerations*

Before substantial resources are devoted to a study, the possible results the study could yield need to be investigated. In particular, the level of risk that has a high probability of being detected needs to be assessed. In another field, it is becoming increasingly recognized that small clinical trials are usually counterproductive. There is

little probability that they can detect realistic differences in treatment; the only significant differences they can show will almost certainly overestimate the real effect. They might be considered biased against the correct result. The same considerations apply to small epidemiological studies. Studies that have low power of detecting realistic levels of excess risk should not be performed, unless their results can be merged with those from other studies.

Chapter 7 discusses in some detail power calculations for both cohort studies and case-control studies.

(*i*) *The possible role of a case-control approach within a cohort*

The essential feature of a cohort study is that each cohort member is followed from entry into the study to death or to the date at which follow-up ends. There are a number of different approaches, however, to the way in which the information on the relevant exposure variables is collected. Gathering the full information on all cohort members may on occasion be a waste of resources and so prevent more useful activities taking place. Typically, the final comparison will be based on a relatively small group of cases and a much larger group of controls. One can therefore take only a subsample of the controls without affecting appreciably the precision of the comparison. Omitting cases, of course, will lead directly to a loss in precision. The questions to face in the design, therefore, are whether and how one can limit the number of individuals for whom full information is obtained without jeopardizing either the validity or the precision of the study. The problem of precision, and of how the sampling might actually be performed, is discussed in detail in Chapters 5 and 7, where different sampling schemes are considered. The sample might consist, for example, of sets of controls, each set individually matched to a particular case, or it might consist of an unstructured subcohort (one in ten of all individuals, say). Designs in which a subcohort is chosen at the start of the study to constitute the control group are discussed by Prentice (1986). Here we consider the question of validity. The main options open to an investigator, set out in Table 1.14, are, first, to wait until the deaths (or other events) of interest have occurred, and then to obtain the information from only a sample of the rest of the cohort; second, to obtain the information on the entire cohort but process it only on a sample; thirdly, to obtain and to process information on the entire cohort, but to use only a sample of the entire cohort together with all the deaths of interest for the statistical analysis.

When the investigator can choose his approach, as would be the case in a prospective study, the design should specify at the start of the study for each variable under investigation the time in the study when information or samples are to be obtained, and when assay or processing is to be performed. The aim should be to reduce the overall burden of data collection and laboratory assays to the minimum consistent with validity, so that attention can be focused on maximizing the quality of the information obtained.

In historical cohort studies the investigator would usually not have the choice. He or she would simply have to decide whether further information, such as smoking histories, was worth obtaining retrospectively, and there would seldom be much value

Table 1.14 Possible approaches to data acquisition

Alternative approaches to the acquisition and treatment of exposure variables	Implication, given information required on all deaths (from a given cause)	Examples and remarks
1. Information obtained only on a sample of the cohort	Information collected when death status is known. Retrospective data must be equivalent to prospective data, and strictly comparable between cases and controls. Neither death nor disease state should affect the variable being measured.	Suitable for genetic markers, or when information from other sources is available independent of the study. No variable in which recall bias may operate should be ascertained in this manner, nor any metabolic or immunological marker affected by the disease in question.
2. Information (or biological specimens) collected on all the cohort, but processed (or assayed) only on a sample	Processing or assays performed only when death status is known. Long-term storage of unassayed samples required	Method of choice when the assay or processing makes heavy demands on resources (e.g., processing seven-day dietary diaries, assaying most metabolic and immunological markers). Essential to demonstrate that storage does not invalidate assay, and that records can be stored safely
3. All information available on entire cohort, but statistical analysis uses only a sample	None	A useful approach to exploratory data analysis (see Chapter 5)

in obtaining it on the whole cohort. He might also decide that further information was required to clarify the results for just one, or a few, causes of death. For example, in a study of Danish brewery workers, known to drink large quantities of beer, Jensen (1979) found an excess of oesophageal cancer (and of course many other diseases). An important question in the epidemiology of oesophageal cancer is whether the exposure of importance is alcohol itself or a particular type of alcoholic drink. In this situation, was it the beer that caused the excess, or were the oesophageal cancer cases heavy consumers of other types of alcohol as well? A subsequent case-control study showed that the association was mainly with beer, i.e., that in this cohort of brewery workers, heavy beer drinking was sufficient to lead to an excess of oesophageal cancer (Adelhardt *et al.*, 1985).

1.5 Interpretation

The initial aim of most epidemiological studies is to determine, to the extent possible with the available data, whether some exposure represents a carcinogenic hazard. The previous section attempted to define what data should be collected for this purpose; this section considers how these data may be used. Criteria for assessing whether an observed association is likely to be causal were discussed at length in Chapter 3 of

Volume 1. Increasingly, however, the demand is not for qualitative evidence of carcinogenicity, but for quantification of the degree of risk. The analysis of the results of a cohort study should aim to extract the maximum quantitative information that the data can yield. The extent to which an analysis produces coherent quantitative descriptions of risk depends at least in part on the manner in which the exposure data are handled – what composite measures of exposure are used, for example.

It should be borne in mind that cohort studies are often the result of an unusual or even unique opportunity. Prospective studies are rarely undertaken, because of their cost and duration. Historical studies are often focused on one of the few cohorts that may exist for which there is clear evidence of exposure to the agent of interest. Either way, there may be few opportunities to repeat the study, unlike retrospective case-control studies, where, for most major sites, a large number of studies have been performed. There is thus an added onus on the investigator to exploit his material to the full.

The two major aspects of an observed excess risk that merit attention are indicators of a dose-response and the evolution of risk with time.

(*a*) *Dose-response relationships*

Both to identify the groups at highest risk and to demonstrate a dose-response, categorization of exposure is helpful, even if no reliable measure of exposure levels is available. Job categories, for example, can be classified as low, moderate or high exposure, as in the Montana smelter workers study (Chapters 4 and 5). The data themselves may indicate that some jobs comport a particularly high risk, as in the South Wales nickel workers (Chapters 4 and 6), although one must be wary of circular arguments. In some instances, little information may be available to classify either job categories or individuals. In this situation, length of employment may provide the best measure of degree of exposure. As mentioned earlier, time variables associated with exposure should be accurately recorded in a cohort study. Given the large effect that errors of measurement can have on estimates of the maximum degree of risk, and of the shape of a dose-response, the accuracy with which time variables are recorded may often make them more valuable than less accurate measures of level of exposure in distinguishing risk, even if the latter might appear *a priori* to be more relevant.

The exposure information available for the six studies summarized in Appendices IA–IF is shown in Table 1.12. When adequate dose information is available, one has to decide how to incorporate it most informatively into the analysis. In Chapter 4, straightforward methods of analysis are described in which categorization of an exposure history into a few levels is required. Chapters 5 and 6 discuss how continuous exposure levels can be treated in the analysis. In each chapter, the aim is the same: how can the data be best utilized to assess causality and to throw quantitative relationships into the clearest light.

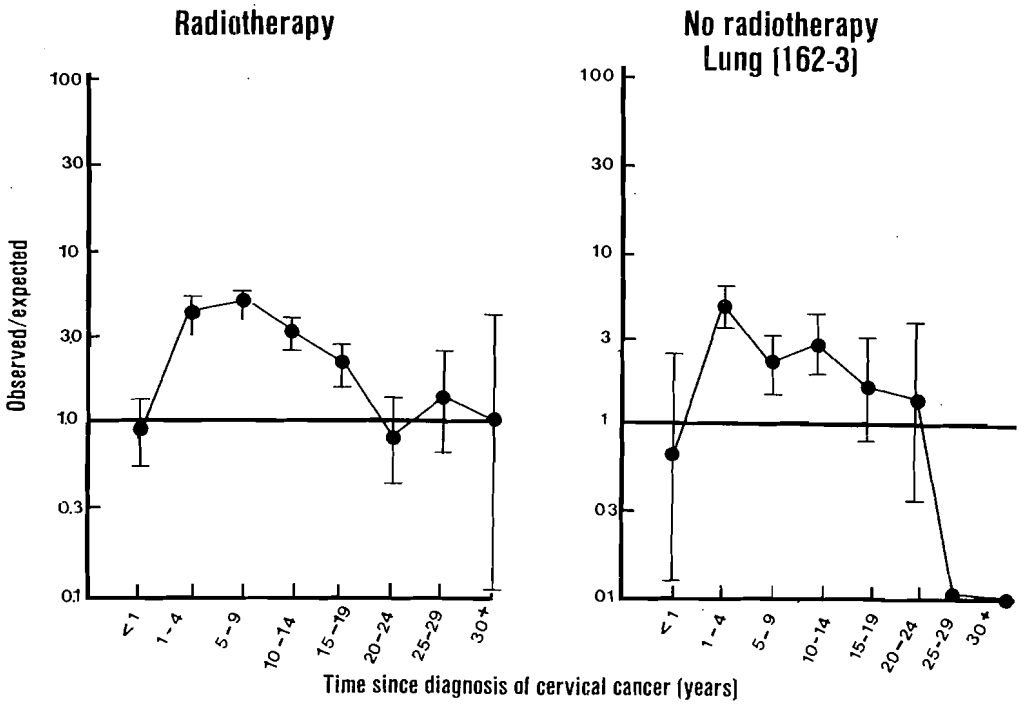
(*b*) *Time relationships*

In describing the excess risk associated with an exposure, it is of interest to know not only the level of risk that can be expected, but also when that excess is likely to occur.

Among continuing cigarette smokers, lung cancer incidence rises with the fourth power of duration of smoking. Mesothelioma rates rise as the third or fourth power of time since first exposure to asbestos. Excess leukaemia mortality forms a wave with a peak at some five years after short-term exposure to radiation. These relationships, discussed in greater detail in Chapter 6, indicate that the temporal behaviour of excess cancer incidence following carcinogenic exposures exhibits well-defined patterns. These patterns may vary from site to site and between exposures, but the pattern of the change with time of an observed excess risk can be determinant in deciding whether an association is causal.

An example is given by a follow-up study of women treated by radiation for cancer of the cervix, in which a large (fourfold) excess of lung cancer was observed (Day & Boice, 1983). Luckily, a nonirradiated group was also included in the study, in whom a similar excess was seen, so that the excess was clearly independent of the irradiation. It is instructive to examine, however, the evolution after the radiation treatment of the excess lung cancer risk, displayed in Figure 1.3. The change in risk with time is unlike that seen for lung cancer in other studies of radiation, in which a genuine exposure-related excess was observed. The normal pattern is for the excess to appear only some ten years after exposure starts. When examined by age at irradiation (i.e., diagnosis of cervical cancer), the picture among those under 50 years of age at irradiation is even more extreme than that shown in Figure 1.3, and for women under age 40 at irradiation the initial excess was nearly 20-fold. Thus, even without the

Fig. 1.3 Observed to expected ratios of lung cancer by time since diagnosis of cervical cancer for patients with invasive cervical cancer treated with radiotherapy and patients with invasive cancer not treated with radiotherapy; 80% confidence intervals presented. From Day *et al.* (1983)



evidence from the nonirradiated group, the shape of the time-risk curve is such that one would feel confident that the observed excess risk is not causally related to the radiation. An alternative explanation, that the excess lung cancers observed are in fact misclassified metastases from the original cervical cancer, fits the observed changes with time and age closely. Smoking, an obvious confounder, is responsible for only a small part of the excess (Day *et al.*, 1983).

Apart from the evolution of risk with time since the start of exposure and with duration of exposure, the change in risk after exposure stops is also of importance. Not only does it aid interpretation, in that a decreasing excess risk after exposure stops would be further evidence of a causal relationship, but the effect of removing exposure is of major intrinsic interest. It is the main epidemiological guide to the effects of intervention measures, and to when public health measures may yield results. Part of the analysis of a cohort study should be orientated specifically at this aspect, and in the design of the study particular efforts should be made to include those formerly exposed.

(c) *Problems in the interpretation of cohort studies*

A number of issues arise in the interpretation of cohort studies, some of which are due to the longitudinal nature of the data acquisition, some of which are common to most analytical studies in which emphasis is put on quantification.

(i) *Choice of comparison groups and the healthy worker effect*

For studies in which subcohorts can be distinguished in terms of level or duration of exposure, most weight in the interpretation will usually be given to comparisons between subgroups within the cohort. Internal comparisons may not always be possible, however, and reliance may have to be put on comparisons with population rates external to the cohort. The question is then to decide which rates to use. Industrial cohorts usually live in urban areas; manual workers smoke more than professional and managerial groups; for a variety of reasons national rates may be inappropriate. Under these circumstances, one can attempt to use rates for a specific socioeconomic group, or for a locality if these are available. The issue of which rates to use to calculate expected numbers is well discussed in a report of the UK Medical Research Council (MRC Environmental Epidemiology Unit, 1984).

A common experience when studying cohorts of employed individuals is that the risk of dying in the first years after entry into the cohort, i.e., after identification as an employed individual, is less than that of the general population. Fox and Goldblatt (1982) have shown conversely, using UK census data, that mortality among the unemployed is particularly high. The reduction in mortality, the healthy worker effect, varies between disease categories and appears to be smaller for cancer than for other major groups. For cancer, the effect also appears to be smaller for cancer incidence than for cancer mortality, reflecting the fact that those with cancer are more likely to have left their job. The healthy worker effect tapers off as years pass since entry into the cohort, unfortunately confounding any real increase in risk with years since first exposure.

An analogue of the healthy worker effect may also occur when the cohort is made up of those who respond to invitations or mailed questionnaires. In the British doctors study, those who replied to the initial questionnaire had, in the first few years of follow-up, an overall mortality considerably less than that of all British doctors. As Doll and Hill put it: '. . . there may be some general association between mortality and the tendency not to reply to such an enquiry, whether the tendency be due to a deliberate refusal (which is rare) or a mere neglect of things (which is frequent). In this respect it is perhaps not too fanciful to note that one non-replier died of smallpox and another of diabetic coma.' In the controlled trial of breast cancer screening in New York, among those invited to screening, the women who accepted had half the overall mortality of those who did not attend, even though they were at considerably higher risk for breast cancer (Shapiro *et al.*, 1982).

The healthy worker effect, since it produces lower mortality rates for many causes of death, may mask real effects. This masking is particularly difficult to interpret if comparisons are made with an external standard population. The more that comparisons are made between different exposure categories within the cohort, the less distortion to the overall interpretation will be caused by the healthy worker effect.

One aspect of the healthy worker effect requires special treatment since it is not eliminated by confining comparisons to those between subgroups of the cohort. Employment status often changes due to ill health. People may retire because they are chronically sick or, because of incapacity, move to lighter work or change jobs. Mortality is therefore likely to be particularly high in the year or two succeeding changes in employment, and conversely relative low in those not changing employment. Odd peaks and troughs may thus appear if risk is examined in relation to time since, or time before, change in job category. One means commonly used to alleviate this problem is to lag changes in status by a number of years – often two or three. In this way, the first year or two after retirement, and the deaths that occur within them, are treated as if the individual were still employed. The matter is discussed further in Chapter 3 (p. 87).

(ii) *Losses to follow-up*

The validity of a cohort study depends fundamentally on complete ascertainment of the events of interest (e.g., cancer deaths) and correct computation of the population at risk. Unless at the start of the study one can be confident that losses to follow-up can be limited to the levels seen in Table 1.10, a study should probably not be launched.

Individuals leave the population at risk either through death, or through migration to a country or region where the follow-up mechanisms of the study are not operative. If an individual has left the population at risk, i.e., the observable cohort, but this fact is unrecorded, then he will continue to contribute to the person–years at risk, but can no longer contribute to the events of interest. Mortality and incidence rates for each cause will be biased downwards. An evaluation of the extent to which follow-up losses have occurred is important, documentation of low loss rates adding to the credibility of the results. Thus, at the date chosen as the end of follow-up, the status should be ascertained of those still thought to be active members of the cohort and under

observation. The proportion not found gives the proportion lost to follow-up. In some situations, this ascertainment may be laborious, and it might be undertaken on a sample, if selected in unbiased fashion.

(iii) *Biases due to errors of measurement*

One of the advantages of cohort studies over case-control studies is that information on exposure is obtained before disease status is ascertained. One can therefore have considerable confidence that errors in measurement are the same for individuals who become cases of the disease of interest, and the remainder of the cohort. The complexities possible in retrospective case-control studies because of differences in recall between cases and controls do not apply. Measurement error will bias estimates of relative risk and standardized mortality ratio (SMR); the extent of the bias is indicated in Table 1.15 for different rates of misclassification.

The shape of dose-response curves, and not just their overall level, will also be altered by error in measurement (Doll & Peto, 1978). A linear dose-response, for example, may be transformed into one concave upwards, concave downwards, or

Table 1.15 Bias in the estimation of the odds ratio associated with a dichotomous exposure variable in a case-control study if there is misclassification of exposure levels. The body of the table shows the ratio of the odds ratio estimated using misclassified data to the true odds ratio[a]

P	β	$R_E = 2$						$R_E = 10$					
		$\alpha = 0$	0.1	0.2	0.3	0.4	0.5	$\alpha = 0$	0.1	0.2	0.3	0.4	0.5
0.1	0.0	1.00	0.99	0.98	0.97	0.96	0.95	1.00	0.91	0.84	0.77	0.72	0.68
	0.1	0.76	0.74	0.72	0.69	0.67	0.64	0.57	0.50	0.43	0.37	0.32	0.28
	0.2	0.68	0.66	0.64	0.62	0.59	0.57	0.42	0.36	0.30	0.26	0.22	0.19
	0.3	0.64	0.62	0.60	0.58	0.56	0.54	0.34	0.28	0.24	0.20	0.17	0.14
	0.4	0.61	0.59	0.57	0.55	0.53	0.52	0.30	0.24	0.20	0.17	0.14	0.12
	0.5	0.59	0.57	0.55	0.53	0.52	0.50	0.26	0.21	0.17	0.14	0.12	0.10
0.5	0.0	1.00	0.92	0.86	0.81	0.78	0.75	1.00	0.55	0.40	0.32	0.28	0.25
	0.1	0.95	0.86	0.80	0.75	0.71	0.68	0.92	0.48	0.34	0.27	0.23	0.20
	0.2	0.92	0.82	0.75	0.70	0.66	0.62	0.85	0.42	0.29	0.23	0.19	0.17
	0.3	0.88	0.78	0.71	0.65	0.61	0.57	0.79	0.36	0.25	0.20	0.16	0.14
	0.4	0.86	0.74	0.67	0.61	0.57	0.53	0.74	0.32	0.22	0.17	0.14	0.12
	0.5	0.83	0.70	0.63	0.58	0.53	0.50	0.70	0.27	0.18	0.14	0.12	0.10
0.9	0.0	1.00	0.68	0.61	0.58	0.56	0.55	1.00	0.19	0.15	0.13	0.12	0.12
	0.1	0.99	0.66	0.60	0.57	0.55	0.54	0.99	0.18	0.14	0.13	0.12	0.12
	0.2	0.99	0.65	0.58	0.56	0.54	0.53	0.98	0.17	0.13	0.12	0.12	0.11
	0.3	0.98	0.63	0.57	0.54	0.53	0.52	0.97	0.16	0.13	0.12	0.11	0.11
	0.4	0.98	0.61	0.56	0.53	0.52	0.51	0.96	0.15	0.12	0.11	0.11	0.10
	0.5	0.97	0.59	0.54	0.52	0.51	0.50	0.95	0.14	0.12	0.11	0.10	0.10

[a] From Tzonou *et al.* (1986)
R_E = true odds ratio between exposure E and disease
P = (true) proportion of population exposed to E
α = proportion of those truly $E+$ classified as $E-$
β = proportion of those truly $E-$ classified as $E+$

Table 1.16 Effects of misclassification on the shape of a linear dose-response curve – exposure grouped into three categories

	Real exposure		
	Low	Medium	High
Misclassification matrix			
Observed Low	0.65	0.175	0
exposure Medium	0.35	0.65	0.35
Hlgh	0	0.175	0.65
Relative risk of real exposure	1.0	5.0	9.0
(*a*) Creation of a dose-response curvilinear *upwards*			
Real distribution of:			
Control population	76%	16%	8%
Case population	33%	34%	33%
Observed relative risk	1.0	2.2	6.3
(*b*) Creation of a dose-response curvilinear *downwards*			
Control population	33%	34%	33%
Case population	6.7%	33.3%	60%
Observed relative risk	1.0	4.4	4.4

linear with lower slope. What happens will depend on the distribution of the exposure in the cohort under study, and on the misclassification rates. Examples are given in Table 1.16.

Interpretation of the results will be sharpened if information is available on the misclassification rates, and, in this respect again, cohort studies, particularly prospective studies, have a clear advantage over case-control studies. Contact with the study cohort during the period of follow-up will permit assessment not only of the change in exposure variables during this period, but also estimation of the misclassification rates. These rates are then equally applicable to future cases and to the controls, and valid adjustments can be made to the observed relative risk and dose-response curves and also to the corresponding confidence intervals. Furthermore, correct use can be made of information on confounding variables (Clayton, 1985; Clayton & Kaldor, 1987). One should note, however, that repeating the same questionnaire may well not provide a second, independent observation; it may simply repeat the same errors. A more subtle approach will often be required. In case-control studies, repeat measurements are rarely available and, if at all, only on the controls. Their applicability to the cases will be questionable.

(iv) *Lack of information on confounding factors*

In an earlier section, we considered how information on confounding variables might be acquired. In many historical cohort studies, however, no opportunity will exist for further data collection, and one is left with the problem of interpreting the results, knowing that information on an important variable is missing. The way in which a

Table 1.17 Excess relative risk of lung cancer among cervical cancer cases

	Proportion in the population	Relative risk of lung cancer	Average relative risk for lung cancer
Cervical cancer patients	Smokers 71% Nonsmokers 29%	10 1	$(10 \times 0.71) + (1 \times 0.29) = 7.4$
Women in the general population	Smokers 34% Nonsmokers 66%	10 1	$(10 \times 0.34) + (1 \times 0.66) = 4.1$

quantitative assessment can be made of the effect such factors may have had can be illustrated by an example from the study of second cancers among patients diagnosed with an initial cancer of the cervix (see §1.5*b*) (Day *et al.*, 1983). The study cohort contained 84 000 women diagnosed with in-situ lesions, among whom distant metastases would be infrequent. Among these women, lung cancer rates were increased more than two fold (SMR = 2.1). The question arises as to whether this excess could be due to smoking, known to be more frequent among cervical cancer patients from other studies. No information was available on the smoking history of the study cohort, since cancer registry records were the only source of information. One can, however, calculate approximately the excess risk likely to be due to smoking using data from other sources. In a study by Buckley *et al.* (1981), 71% of cervical cancer patients had never smoked compared to 34% of controls. Assuming that the relative risk of lung cancer among smokers compared to nonsmokers is tenfold, one can evaluate the excess relative risk of lung cancer among cervical cancer cases as outlined in Chapter 2 of Volume 1 and shown in Table 1.17. The predicted value of $(7.4/4.1) = 1.8$ is close to the observed value of 2.1 – certainly within the bounds of statistical error – and one would feel confident that smoking was a satisfactory explanation of the observed lung cancer excess. By contrast, among the 96 000 patients with invasive cervical cancer, in the first ten years of follow-up the excess risk of lung cancer was about fourfold (SMR = 3.9). This excess is clearly considerably too large to be explicable in terms of smoking.

Thus, use of concomitant information, even on populations distinct from the cohort under study, can remove much of the uncertainty due to unrecorded confounding factors.

(v) *Multiple comparisons*

In most cohort studies, an assessment will be made of a large number of disease categories as endpoints: there may be more than 30 sites of cancer for which observed and expected values are compared. Some of these comparisons can be anticipated to achieve nominal significance levels just by chance. The situation is often compounded by the inclusion in the study of more than one factor of interest, and in the analysis by using a variety of ways of examining each factor. Search for interaction effects, or

looking at subgroups defined by several variables simultaneously, will increase further the possible comparisons.

The topic of multiple comparisons was considered in Volume 1, in the context of case-control studies, where there may be many exposure variables. In cohort studies, the multiplicity of the comparison refers more often to disease categories. Most cohort studies, however, are not launched in an intellectual vacuum. Animal experiments may suggest the site of action of the exposure; the route of administration may indicate which sites are most exposed; preliminary data, from official statistics or proportional mortality studies, may have drawn attention to a particular risk. There will therefore be one, or a few, sites at which any effect might be hypothesized to occur. Results for these sites should be interpreted differently from results for other cancers, which should probably be regarded as hypothesis generating, and the significance values modified accordingly, for example by multiplying by the number of such sites. For the former sites, a stricter interpretation in terms of hypothesis testing would be appropriate. Feinleib and Detels (1985) refer to the reporting of results, nominally significant but outside the original aim of the study, as 'post-hoc bias'.

(vi) *Identification of forerunners of disease, rather than causes*

We have stressed one powerful feature of cohort studies, that measures of physiological status can be made before the appearance of disease clouds the picture. Care must be taken, however, to ensure that the levels of a particular metabolic parameter have not been influenced by a preclinical disease state. An association that appears to be causal may be a reflection of an early state of disease. A number of reports in the late 1970s and early 1980s indicated that serum cholesterol levels were low in individuals who subsequently developed cancer. One interpretation was that low cholesterol levels predispose to cancer. An alternative was that in the year or two immediately preceding clinical onset, low cholesterol levels may be the result of early disease (Rose & Shipley, 1980). Opinion now favours the second of these explanations, since low levels are seen only immediately before clinical onset, and not five or more years before. A review is given by McMichael *et al.* (1984). A similar fate seems to have overtaken earlier reports of low retinol levels observed before cancer onset (see Wald *et al.*, 1980; Kark *et al.*, 1981). The lesson to be drawn is that the association must be examined in relation to the time between measurement of the parameter and disease onset. If the association weakens steadily as the time interval increases, serious doubt would be cast on its interpretation as causal. If the association remains strong as the interval increases, however, one would favour a causal explanation. This behaviour is seen, for example, in the association of hepatitis B and liver cancer (Beasley *et al.*, 1981) and the association of the Epstein–Barr virus and Burkitt's lymphoma (de-Thé *et al.*, 1978).

(vii) *Conclusions to be drawn from negative results*

Much emphasis is put on criteria for interpreting positive results, the extent to which they can be taken as indicating causality, and on the degree to which they provide

quantitative measures of excess risk. For studies in which no excess risk is demonstrated, a complementary approach should be taken. The data should be examined for their adequacy in ruling out a positive effect and for the level of excess risk with which they are compatible, and also for whether alternative explanations are possible, i.e., whether bias or confounding may have produced an apparently negative result when a real effect existed. The evaluation of apparently negative evidence has been the topic of a recent publication (Wald & Doll, 1985). The following points are among those that should receive attention.

- What are the confidence limits for the excess risk? In Table 2.10, values are given for confidence intervals for the SMR. Clearly, even a moderately sized study, with, say, 50 events of interest, with an estimated SMR of exactly 1.0, cannot exclude increases in risk of the order of 30%.
- How do the dose levels observed in the present study compare with the levels to which other segments of the population are exposed?
- Had sufficient time elapsed between the start of exposure and the end of follow-up for a potential risk to have expressed itself fully? In this respect, it is useful to examine the excess risk seen ten years or more after first exposure, for which the confidence intervals will usually be considerably wider than for the cohort overall.
- Is there any reason to suspect that the cohort is at substantially lower risk than the general population? In a study of nuns in the United Kingdom, for example (Kinlen, 1982), one would expect particularly low rates for cancer of the oesophagus since they use neither alcohol nor tobacco. The observed excess (11 observed against 5.57 expected), which numerically would be of marginal interest in most circumstances, is therefore of particular note.
- What is the consistency with other studies?

1.6 Proportional mortality studies

An extreme example of loss to follow-up occurs when one has no accurate data on the composition of the cohort, but one has a set of death records. The proportion of deaths due to each cause arising from a particular cohort is known, but not the absolute mortality rates. One is then led to a study of proportional mortality rates, comparison being made either with the proportions seen in the general population or among subgroups in the study group. A similar situation arises in cancer registries, where one may have, for example, information on occupation, but obtained in a way not comparable to available census data. One then has a proportional incidence study.

Proportional studies have been used considerably in descriptive cancer epidemiology where, in the absence of corresponding census data, they may draw attention to unusual or contrasting patterns of cancer occurrence (Parkin, 1986). In analytical epidemiology, proportional mortality studies may be of considerable value in the initial stages of an investigation. They may indicate a fruitful orientation for later, more rigorous studies, and certainly provide a cheap and rapid way of taking an initial look at a set of data. Results of studies of proportional mortality rates suffer from a particular difficulty in interpretation, in that a proportionate excess can reflect either an

excess in the absolute rate for that disease, or a deficit in the absolute rates for some of the other causes. It is unlikely, however, that large proportionate excess rates would be produced in this way. The approach is formally equivalent to a case-control study based on deaths, in which the cases have died from one cause of death and the controls are selected from deaths from all other causes. Seriously biased results can be obtained, as for example an apparent strongly protective effect for cigarette smoking against dying from mesothelioma (Blot *et al.*, 1980); but, provided one is aware of the dangers, useful results can be obtained.

One must be particularly careful in conducting proportional mortality studies to include all deaths, or at least the great majority. Differential exclusion of a particular cause of death not only decreases the proportional rate for that disease, but increases the proportional rate for all others. Initial reports of an excess risk based on high proportional rates from incomplete information may well be modified when more complete data are available, as happened with claims of an excess of leukaemia in Portsmouth (USA) shipyard workers (see Example 3.7). The problem is compounded by the fact that in a full cohort study the extent to which follow-up is incomplete can be explicitly stated (as in Table 1.9), whereas, in a proportional study, almost by definition, the degree of incompleteness is unknown. If it were known, it would imply that sufficient information on the cohort and the follow-up would be available for a full cohort analysis to be performed (i.e., calculation of rates rather than proportions).