Chapter 1
Introduction to cancer epidemiology

Interest in cancer has grown during the past century as infectious diseases have increasingly been controlled as the result of improved sanitation, vaccination and antibiotics. Although this interest is relatively recent, cancer is not a new disease. Autopsies of ancient Egyptian mummies have shown the presence of bone tumours and possibly other neoplasias (Brothwell, 1967). Symptoms of what can be assumed to be malignant diseases were also described in Chinese and Arabic medical writings. By the time of Hippocrates in the 4th century BC, many types of tumour were clinically recognized and described. Hippocrates introduced the term carcinoma from the Greek word *karkinos*, for crab: he saw cancer as crab-like in its spread through the body and in its persistence (Long, 1928). Some 600 years later, Galen distinguished three types of tumour: 'tumours according to nature', which included all normal physiological swellings, such as enlargement of the breast with normal female maturation; 'tumours exceeding nature', which included the productive process following injury, such as the proliferation of bone that occurs during the reuniting of a fracture; and 'tumours contrary to nature', which included what we now define as neoplastic growths, as well as many inflammatory lesions (Long, 1928).

However, it was not until the end of the 18th century that cancer began to be studied systematically and intensively. Bichat (1771–1802) described the pathology of many neoplasms in humans and suggested that cancer was an 'accidental formation' of tissue built up in the same manner as any other part of the organism. Some decades later, Müller (1801–58) and Virchow (1821–1902) extended Bichat's findings, using the microscope to show that cancerous tissue was made up of cells (Long, 1928).

Ever since, pathologists and clinicians have considered cancers in the various organs of the body as being in many respects completely different diseases with distinct morphologies, clinical manifestations and prognoses. But only during the past few decades has it emerged that their causes also differ enormously. As a discipline, epidemiology has been crucially important in defining the causes of different cancers and in evaluating preventive measures.

1.1 What is cancer epidemiology?

A recent definition of epidemiology is given in the dictionary compiled by Last (1995):

> "Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems."

Cancer epidemiology is the branch of epidemiology concerned with the disease cancer. Therefore, this definition is as valid to cancer epidemiology as it is to epidemiology in general.

The first thing to note in the above definition is that epidemiology is concerned with events that occur in populations: the primary units of concern are groups of people, not separate individuals. This is what differentiates epidemiology from clinical medicine. Epidemiological studies are concerned not only with people who get a disease, but also with those who do not, and in particular how these two groups may differ. Thus, unlike a doctor in clinical practice, who is usually concerned only with patients, the epidemiologist is concerned equally with the population from which the patients came. Whereas clinicians direct their questions towards a particular patient—'What is wrong with this patient? What treatment is appropriate?'—the epidemiologist asks similar questions about whole communities—'What are the leading causes of death or disability in this population? What can be done to reduce them?' Thinking in epidemiological terms often seems foreign to clinicians and other health-care professionals, who are used to thinking of the problems of each individual patient.

While it seems obvious that cancer epidemiology should focus on the disease known as cancer, this is not necessarily the case. In fact, cancer epidemiological studies may focus on precursors of cancer; for example, on cervical intraepithelial neoplasia (CIN) as a precursor of invasive cervical cancer, or on chronic atrophic gastritis as a precursor of stomach cancer. Alternatively, the object of the study may be a characteristic that is related to cancer, such as growth or fertility. Thus, epidemiologists are concerned not only with illness, disability and death, but also with health and with preventing disease.

The epidemiologist is interested in the distribution of disease in a population. Which types of people are at a higher risk? How does the disease frequency change over time? How much does it vary from place to place? These are summarized as 'Who?', 'When?' and 'Where?'. The most basic task of cancer epidemiology is to describe the occurrence of human cancer, noting differences, for example, between males and females, between people of different ages, between socioeconomic classes, between occupations, between time periods, between areas of a country, and between countries. These observations are the starting points in epidemiological research. Like a detective, the epidemiologist then pursues the most promising clues.

Epidemiology aims to answer not only 'Who has what, when, and where?', but also to find out why. Why are some at higher risk than others? Are associations between certain factors and increased risk of disease

causal or spurious? Epidemiologists seek to address such basic issues, viewing their work as the search for a relationship between two factors. The first, called exposure (or risk factor), refers to any factor that can affect a person's health: such factors include environmental agents, such as sunlight, air pollution and occupational exposures; lifestyle variables like diet, smoking and physical exercise; and constitutional factors, such as blood type and other genetic traits. The second is the disease itself (or other health-related event of interest). The epidemiologist must examine whether there is an association between an exposure and a disease, and decide whether the observed relationship is likely to be causal. Questions to be addressed might be 'Is lung cancer associated with cigarette smoking?' or 'Does cigarette smoking cause the disease?'.

Epidemiology is the only source of direct scientific evidence about exposure effects and the preventability of disease within human populations. Laboratory scientists have identified carcinogenic compounds in tobacco smoke and have been able to produce respiratory cancers in experimental animals by forcing them to inhale cigarette smoke (IARC, 1986). However, the argument that cigarette smoking causes lung cancer in humans would remain unconvincing if epidemiologists had not also demonstrated that lung cancer occurs much more often in smokers than in non-smokers. But, in contrast to the laboratory-based sciences, the strategy in epidemiology is usually to observe and compare, rather than to experiment, as major ethical and practical considerations limit the possibilities for experimental studies upon humans.

In addition to establishing whether particular exposure–disease associations exist, epidemiology attempts to measure their strength. Thus, after finding an association, the next stage is always to determine the magnitude of the possible effects. To do so, epidemiologists ask questions such as 'By how much does cigarette smoking increase the risk of lung cancer?' or 'By how much does hepatitis B vaccination reduce the risk of liver cancer?'.

One of the most important roles of epidemiology is to learn about the causes of disease, or its natural history, knowledge that can lead to the introduction of preventive measures. Even when the biology of a disease is not fully understood, epidemiology can identify a cause and a means of prevention. Complete knowledge of causal mechanisms is not essential for effective preventive strategies, although in the spirit of scientific enquiry accepted in western culture, knowledge of the mechanisms involved is desirable in itself. We can greatly reduce our risk of developing lung cancer if we do not smoke cigarettes, without knowing what specific component of cigarettes is carcinogenic, nor precisely how these substances affect the control of cell growth.

The epidemiologist is to society what the doctor is to the patient, and epidemiology is thus part of the foundations of public health. Epidemiology describes and measures the occurrence of disease in the community, so that questions can be addressed such as 'Is the disease (or

condition) a priority for health intervention?' and 'Why do cases occur and why does the condition persist in the community?'. Epidemiology also helps in choosing the health programmes that are most likely to control a particular health problem and in evaluating their impact in the community.

1.2 A brief history of cancer epidemiology

Epidemiology has its origins in the idea present in many ancient civilizations that environmental factors can influence the occurrence of disease. This idea was clearly expressed about 2400 years ago by Hippocrates:

> "Whoever wishes to pursue properly the science of medicine must proceed thus. First he ought to consider what effects each season of the year can produce; for the seasons are not all alike, but differ widely both in themselves and at their changes. The next point is the hot winds and the cold, especially those that are universal, but also those that are peculiar to each particular region. He must also consider the properties of the waters (...) and how the natives are off for water, whether they use marshy, soft waters, or such as are hard and come from rocky heights, or brackish and harsh. The soil too, whether bare and dry or wooded and watered, hollow and hot or high and cold. The mode of life also of the inhabitants that is pleasing to them, whether they are heavy drinkers, taking lunch, and inactive, or athletic, industrious, eating much and drinking little.(...) Using this evidence he must examine the several problems that arise. For if a physician know these things well, by preference all of them, but at any rate most, he will not, on arrival at a town with which he is unfamiliar, be ignorant of the local diseases, or of the nature of those that commonly prevail; so that he will not be at a loss in the treatment of diseases, or make blunders, as is likely to be the case if he have not this knowledge before he consider his several problems." (Translated into English by Jones, 1923)

Despite his emphasis on the role of environmental factors in causing human disease, Hippocrates believed that cancer was a disease caused by an excess of 'black bile', which was manufactured by both the spleen and stomach, but not the liver. His ideas dominated medical practice during the Middle Ages, and it was only with the advent of the Renaissance that they began to be disputed by a number of physicians. Ramazzini was among these. In his book entitled *De Morbis Artificum* (1713) (translated into English by Wright, 1964), he suggested that the high occurrence of breast cancer among nuns was due to their celibate life, an observation that has withstood the test of time. This observation was confirmed by Rigoni-Stern in a paper published in 1842 (translated into English by De Stavola, 1987).

In 1775, Percival Pott described the relationship between scrotal cancer in chimney sweeps and soot in his *Chirurgical Observations*. His work is

now considered a milestone in epidemiology. Pott was the first to attribute an occupational cause to this disease, raising the possibility of prevention.

> "The fate of these people seems singularly hard; in their early infancy they are most frequently treated with great brutality, and almost starved with cold and hunger; they are thrust up narrow, and sometimes hot chimneys, where they are buried, burned and almost suffocated; and when they get to puberty, become liable to a most noisome, painful and fatal disease." (Pott, 1775)

Pott's work was continued by Henry Butlin (1845–1912), a surgeon at St Bartholomew's Hospital in London. An interesting account of this is given by Waldron (1983):

> "A feature of the disease which exercised those interested in it, and which was connected with its aetiology, was that chimney sweeps' cancer seemed to be almost exclusively an English disease. Cases were virtually unknown on the Continent, in America, or even in Scotland. This problem was fully investigated by Henry Butlin. Butlin took himself to the Continent during the course of his researches and, as the result of meeting and talking to Continental sweeps, he considered that he had established the reasons for the virtual absence of scrotal cancer among them. In part it was due to the protective clothing they wore. As early as 1785, the German sweep is depicted in a close fitting suit complete with head covering [Figure 1.1]. There are no openings for the soot to penetrate nor any loose clothing in which it can lodge. The London sweep, as late as 1851 stands in sharp contrast, a waif-like boy, dirty and in loose smock and trousers. [Figure 1.2]"

Butlin made use of a 'natural experiment' to observe that protective clothing seems to be associated with a reduction in the risk of scrotal cancer.

The development and growth of the field of vital statistics in the 19th century made it possible to study the patterns of cancer occurrence in specific populations. William Farr (1807–83) in England, in collaboration with Marc d'Espine in Geneva, developed a nomenclature system for grouping diseases (Farr, 1975), which formed the basis for the *International Classification of Diseases* (see Appendix 2.2). The adoption of this classification by many countries to code causes of death recorded on death certificates greatly improved the comparability of international mortality statistics. One of the earliest and most extensive reports on international cancer mortality statistics was published by Hoffman in 1915 in a book entitled *The Mortality from Cancer Throughout the World*. A graph from this book is reproduced in Figure 1.3. The first population-based cancer registries, which collect information on all new cases of cancer that occur in a well defined population, were also set up in the first decades of this century (Wagner, 1991) (see Chapter 17).

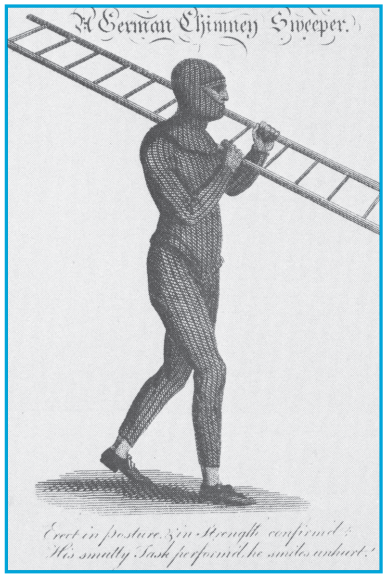

Figure 1.1.
Protective clothing worn by a German sweep in 1785 (reproduced, by permission of BMJ Publishing Group, from Waldron, 1983).

Figure 1.2.
An English chimney sweep in 1851 (reproduced, by permission of BMJ Publishing Group, from Waldron, 1983).

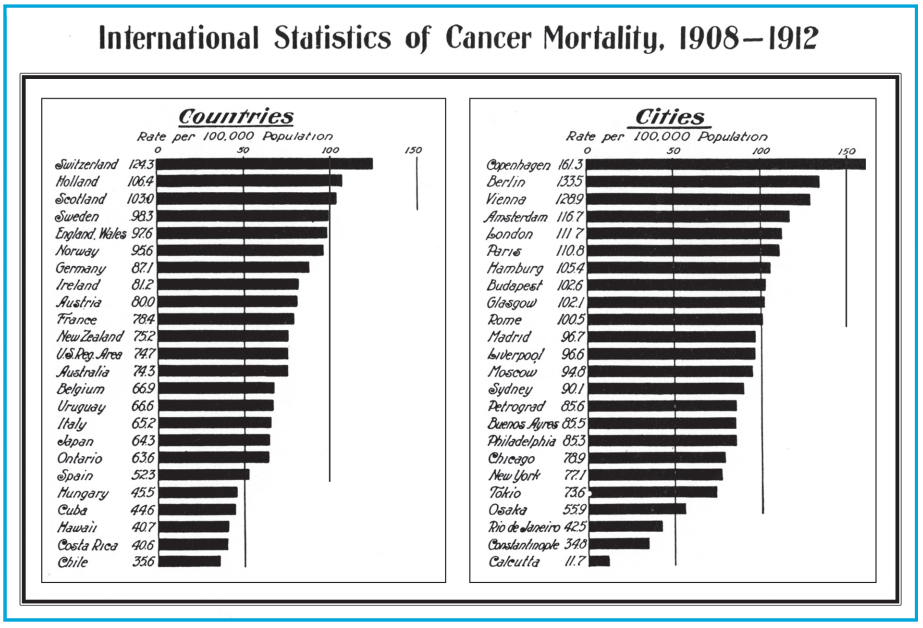

Figure 1.3.
Mortality from all malignant neoplasms for certain countries and cities, 1908–1912 (reproduced from Hoffman, 1915).

Although several studies that would now be considered examples of cancer epidemiology were conducted before the 20th century, the growth of this discipline (and of epidemiology in general) did not begin until the end of the Second World War. In 1954, results from two important cohort studies, now considered as classical, were published. One was the British Doctors' Study (Doll & Hill, 1954), designed to investigate the relationship between tobacco smoking and lung cancer. The other had been set up to determine the risk of bladder cancer in the British chemical industry (Case *et al.*, 1954; Case & Pearson, 1954).

Cohort studies of human populations, particularly of industrial workers, patients treated with radiation and cytotoxic chemotherapy, and victims of nuclear and other disasters, have provided the most convincing evidence of links between exposure to specific agents and cancer.

In cohort studies, the epidemiologist assembles a group of people and collects information to identify those who are exposed to a particular agent (e.g., smokers) and those who are not (e.g., non-smokers). The group is then followed over time, and the disease occurrence (e.g., lung cancer) in exposed individuals is compared with the disease occurrence in the non-exposed group (see Chapter 8). However, the follow-up and observation of disease occurrence in a population is not a simple task. Many diseases, including cancer, occur rarely, so that large numbers of people must be followed up for long periods to observe enough cases with the disease. This requires a great deal of effort from the researchers, cooperation from the study subjects and, above all, a sizeable budget. The high costs and the logistic difficulties involved mean that cohort studies have been favoured in settings where exposure and medical records and vital statistics are carefully collected and available for use, or where society can support the

expense of gathering the necessary information. Despite this, a considerable number of cohort studies have been carried out in developing countries (e.g., Geser *et al.*, 1982; Gupta *et al.*, 1984; Ross *et al.*, 1992). One of the most impressive was the seven-year follow-up of a cohort of 42 000 Ugandan children in an investigation of the etiological role of Epstein–Barr virus in Burkitt's lymphoma (Geser *et al.*, 1982). This study design has also been widely used in developing countries to examine the health effects (including cancer) of exposure to a large number of occupational (Pearce *et al.*, 1994) and other environmental hazards.

The logistic problems and costs encountered with cohort studies have led to the development and increase in popularity of another type of study: the case–control study. In these studies, the epidemiologist compares a group of individuals who have the disease under investigation (termed 'cases'—e.g., persons with lung cancer) with a group without that disease (termed 'controls'), to see whether the groups differ in their past history of exposures (e.g., smoking habits) (see Chapter 9). This type of study is uniquely well suited to studying cancer and other diseases with a long induction period, allowing the epidemiologist to look through time, retrospectively tracing the path back from effect to cause. This is the reverse of the time-sequence adopted in cohort studies. This technique of looking retrospectively is used daily by clinicians when they take case histories. However, case–control studies differ from case series in that they use a control group for drawing comparisons with the group of cases.

One of the earliest studies to make use of the case–control approach was that reported by Broders (1920) on squamous cell epithelioma of the lip in relation to pipe smoking. Unfortunately, Broders failed to describe the method by which he selected the controls. In 1926, Lane-Claypon reported a case–control study of the role of reproductive experience in the etiology of breast cancer. This report is particularly important because it emphasized the need to use a control group for comparison with the cases, and discussed methods for selecting hospital controls to address specific hypotheses. Thereafter until the late 1940s, no further case–control studies of comparable quality were published. There were, however, studies in which the characteristics of cancer patients were compared with those of a group of non-cancer patients; an example, on betel chewing and oral cancers in India, was published by Orr (1933). In 1947, Schrek & Lenowitz reported a case–control study of the relationship between carcinoma of the penis and poor sexual hygiene. In the early 1950s, results from four case–control studies linking cigarette smoking to lung cancer were reported (Schrek *et al.*, 1950; Levin *et al.*, 1950; Wynder & Graham, 1950; Doll & Hill, 1950, 1952), and the following years saw numerous similar studies of many cancers. The 1950s also brought the first methodological developments of the case–control approach (e.g., Cornfield, 1951; Mantel & Haenszel, 1959).

The case–control study has revealed or clarified our understanding of such associations as late first birth and breast cancer (MacMahon *et al.*,

1970); diethylstilbestrol and vaginal clear-cell adenocarcinoma in young women (Herbst *et al.*, 1971); exogenous estrogens and cancer of the endometrium (Ziel & Finkle, 1975; Smith *et al.*, 1975); alcohol and tobacco consumption and cancer of the oesophagus (Tuyns *et al.*, 1977); chronic infection with hepatitis B virus and liver cancer (Prince *et al.*, 1975; Trichopoulos *et al.*, 1978); and human papillomavirus infection and cervical cancer (Muñoz *et al.*, 1992a). The case–control study is particularly appropriate for investigating causal relationships where resources are relatively scarce, such as in developing countries.

A more recent development has been the application of epidemiological principles and methods to the design, conduct and analysis of intervention trials. These are studies in which the exposures being studied are allocated to participants by the investigators themselves (see Chapter 7). This type of controlled experiment has become an integral part of the evaluation of new preventive and therapeutic agents and procedures. For instance, the intervention trial has been used to evaluate the impact of anti-smoking advice on health (Rose & Colwell, 1992), to assess the role of health education programmes in preventing oral cancer in India (Gupta *et al.*, 1986), to evaluate screening programmes for breast cancer (e.g., Shapiro *et al.*, 1971), and to assess the efficacy of hepatitis B vaccination in preventing liver cancer (e.g., Gambia Hepatitis Study Group, 1987).

It is obvious from the low survival from many cancers that if these diseases are to be controlled, we cannot rely solely on increasing the availability of medical care. It is vital to increase our understanding of the genetic, environmental and social factors that foster these diseases, with the aim of applying this knowledge to effective preventive measures. The ultimate goal of cancer epidemiology is to identify risk factors so as to allow the early introduction of effective preventive measures. To achieve this goal, however, cancer epidemiology requires a multidisciplinary approach, bringing together clinicians, laboratory and social scientists, and public health and other health-related professionals.

The contribution of epidemiology to strategies for preventing certain types of cancer was until recently well in advance of our biological knowledge of the disease. However, advances in biology are now providing new tools and directions for epidemiological investigations into the occurrence and distribution of cancer in populations. Incorporation of biological markers into epidemiological research and development of genetic epidemiology are just some examples of this multidisciplinary approach.

Box 1.1. Key issues

- Cancer epidemiology is concerned with the study of the distribution of the disease cancer in populations. Its ultimate goal is to identify risk factors that may lead to early introduction of effective preventive measures.

- Cancer epidemiology, and epidemiology in general, is based on the comparison of groups of people. For ethical reasons, however, epidemiological methods are predominantly observational (i.e., non-experimental). The major challenge for epidemiologists is to identify and make use of 'natural experiments' that will help to answer the question under investigation.

- Cancer epidemiology is a relatively new science, which has matured only in the last half of the 20th century. Despite its youth, it has already contributed greatly to our understanding of the causes of different types of cancers and the evaluation of preventive measures.

Further reading

* Buck *et al.* (1988) compiled and discussed a collection of classic epidemiological papers, including many of those quoted in Section 1.2.