WHO/SDE/OEH/99.7
English only
Distr.: Limited

Environmental Epidemiology: a Textbook on Study Methods and Public Health Applications

Preliminary Edition

World H[illegible] Sust[illegible]elopment and Healthy Environments

United States Environmental Protection Agency

WHO/SDE/OEH/99.7
English
Distr.: Limited

Environmental Epidemiology

A Textbook on Study Methods and Public Health Applications

Preliminary Edition

Editors

Dean Baker
Tord Kjellström
Rebecca Calderon
Harris Pastides

Prepared in Collaboration with:

the United States Environmental Protection Agency

This document was prepared with support from the United States Environmental Protection Agency under cooperative agreement number CR 821793-01-0. Although the contents have benefited from comments from Agency staff, they have not been subjected to formal Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

The illustration of the cover page is extracted from Rescue Mission: Planet Earth, ©Peace Child International 1994; used by permission

Printed in Malta by Interprint Ltd

Environmental Epidemiology

A Textbook on Study Methods and Public Health Applications

Authors and Contributors *

Dean Baker – 3, 5, 6, 8
Bert Brunekreef – 4
Rebecca Calderon – 10, 11
Ruth Etzel – 5
Jack Griffith – 5, 11
Chuan Jie Hong – 8
Barry Johnson – 10, 11
Tord Kjellström – 1, 2, 12
Michal Krzyzanowski – 8
Harris Pastides – 9, 12
Neil Pearce – 3, 7
Göran Pershagen – 2
Carlos Santos-Burgoa – 2
Eugene Schwartz – 5, 8
Colin Soskolne – 2, 8
Elma Torres – 8

* Text was re-arranged during the editing process. The numbers indicate the chapters to which the authors made a substantial contribution.

Acknowledgements

This text was developed under the auspices of the former Office of Global and Integrated Environmental Health of the World Health Organization (WHO). Financial and other assistance was provided by the United States Environmental Protection Agency (US EPA).

Although not formally on the editorial committee, the following individuals made substantial contributions to the editorial process:

Carlos Corvalán
Eugene Schwartz
Colin Soskolne

Jacqueline Sawyer of the Office of Global and Integrated Environmental Health, WHO made a valuable contribution as the technical and language editor.

The following individuals served as reviewers and contributed material:

Donna Baird
Patricia Buffler
Germán Corey
Douglas W. Dockery
Philip Landrigan
Ken Sexton

Work on Chapters 3 and 7 by Neil Pearce was conducted in part during a Senior Research Fellowship of the Health Research Council of New Zealand, and in part during the tenure of a Visiting Scientist Award of the International Agency for Research on Cancer.

The final review and editing was coordinated by Dean Baker of the University of California, Irvine, Center for Occupational and Environmental Health (COEH). The COEH provided valuable assistance to Dr. Baker during the editing process. Stacey Kojaku provided technical assistance for editing and preparing figures.

Table of Contents

Foreword

The objective of this book is to serve as a teaching textbook for a course in environmental epidemiology. The emphasis is on epidemiology study methods – including exposure and health outcomes assessment, study design and implementation, and data analysis – as well as on the application of epidemiology to the management of environmental health hazards.

The original concept was that the book would be an update of the 1983 WHO Criteria Document on Environmental Epidemiology (WHO, 1983). However, it became apparent that an entirely new book was needed because of advances in epidemiological methods and growth in knowledge of environmental health hazards and management since the earlier document was prepared. Several leading researchers and public health practitioners have made contributions to the text as authors or reviewers during the course of preparation.

The intended audience includes persons with some prior training or experience in environmental health or medicine, plus prior training equivalent to a basic epidemiology course and experience in conducting epidemiological studies. The target audience includes academics and researchers in national research institutes in developing countries, but the text has been written so that it should be understandable to public health workers with medical training and experience in epidemiology.

Material from this text and from *Basic Epidemiology* (Beaglehole et al., 1993) has been used in one–week basic environmental epidemiology workshops that have been organized through the World Health Organization's Global Environmental Epidemiology Network (GEENET) in more than 25 countries since 1988 (WHO, 1994). Typical participants in these workshops have included:

- public health and medical officers (at the national or provincial level);
- environmental protection officers or policy makers;
- researchers in environmental health engaged in field studies;
- teachers of public health, environmental health or epidemiology.

Although the text has been developed for use in a one or two week intensive course, some of the chapters have more advanced material that could be introduced, but not covered entirely in a short course. The more advanced material was included so that readers could use the text as a reference when conducting studies. It also means the text would be appropriate for teaching study methods in a standard graduate course in environmental epidemiology.

Final editing and assembling of the text was done by Dr. Dean Baker of the University of California, Irvine. It is anticipated that this text will be revised before final publication as a book. The editorial committee would like to encourage readers to provide comments, suggestions, and technical corrections. Communication can be sent to Dr. Baker at the following:

Professor Dean Baker
University of California, Irvine
Center for Occupational and Environmental Health
Irvine, California 92697-1830 USA
e-mail: **dbaker@uci.edu**

Chapter 1

What is Environmental Epidemiology?

Learning objectives

- Be familiar with the basic ideas used to link environmental hazards to health.
- Be able to describe the historical development of environmental epidemiology and give examples of major environmental health problems for which environmental epidemiology has been instrumental in finding a solution.
- Understand the role of environmental epidemiology in solving environmental health problems.
- Appreciate the background knowledge required to make full use of this text.
- Appreciate the need for training in environmental epidemiology both to carry out environmental epidemiology studies and to interpret environmental epidemiology studies.

1.1 Role of epidemiology in environmental health management

Environment, development and health

During the past thirty years, protection of our environment has become an important public issue. In the 1950s and 1960s, a series of disease outbreaks, including the London Fog incident, Minamata disease and Itai itai disease, that were associated with environmental factors, drew attention to the importance of environmental protection during industrial and urban development. Publication of a number of key texts, such as *Silent Spring* (Carson, 1962), raised awareness of the issue yet higher. In the late 1960s, local and national community action groups formed to promote protection of the environment and ultimately instigated similar activity at the international level. Governments recognized their obligation to respond with visible action and collaborated to hold the UN Conference on the Human Environment in Stockholm in 1972. As one of its contributions to this conference, WHO produced its first comprehensive compendium of the effects of the environment on human health (WHO, 1972).

Since the Stockholm conference, both the need for environmental protection, and awareness of the links between environmental protection and human health, have grown considerably. Major initiatives at the international level have provided new impetus for the protection of human health and environment, and led to the creation of new infrastructure for international collaboration. The United Nations Environment Programme (UNEP), for instance, has been in operation since 1972 and the International Programme for Chemical Safety (IPCS) was established in 1980. More recently, the work of the World Commission on Environment and Development (1987) (the Brundtland Commission) has ensured that the term "sustainable development" has became a familiar concept.

Box 1.1

"Sustainable development" is development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

Source: World Commission on Environment and Development, 1987.

WHO's own increasing involvement in environmental issues was signified by the review of its Commission on Health and Environment (WHO, 1992d) of the impacts of environmental factors on human health. This review was a major contribution to the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in 1992. UNCED produced a blueprint for future action — *Agenda 21* — on environment and development issues (UN, 1992) with a major chapter on protection of human health and reference to health issues made in many of the other 39 chapters. It also initiated a new intergovernmental coordinating structure for such action, namely, the Commission for Sustainable Development (CSD) at the United Nations. WHO is the "task manager" for health issues in the CSD activities to monitor implementation of Agenda 21. A review report on these issues titled "Health and Environment in Sustainable Development" (WHO, 1997) was prepared for the Special Session of the UN General Assembly on the 5-year anniversary of UNCED.

Environmental health management involves reducing the health risks arising from environmental health hazards to the maximum extent possible, for each community or population. In this text, the definitions of hazard and risk are those established by a WHO working group (WHO, 1989). A **hazard** is a source of danger or an agent that can harm human health (a qualitative notion); the **risk** is the probability that health damage will occur (a quantitative notion). To reduce the health risks presented by hazards, we must seek information on:

- the hazards as they occur in particular environments;
- the population groups which are exposed;
- the level to which they are exposed;
- the health impacts that these hazards could or do have;
- potential approaches for reducing exposures;
- acceptability of these approaches in the exposed population (see Figure 1.1).

In this text, environment has been defined broadly as "all that which is external to the human host. It can be divided into physical, biological, social, cultural, etc. any or all of which can influence health status of populations" (Last, 1995).

The environment may be considered as representing one or more of the following settings:

- the home environment: exposure may be determined by, among other factors, personal or family eating habits, cooking facilities, drinking-water and sanitation facilities, type of building (shelter provided), industrial or farming activities carried out in or around the home, hobbies, pesticides applied in the home and garden;

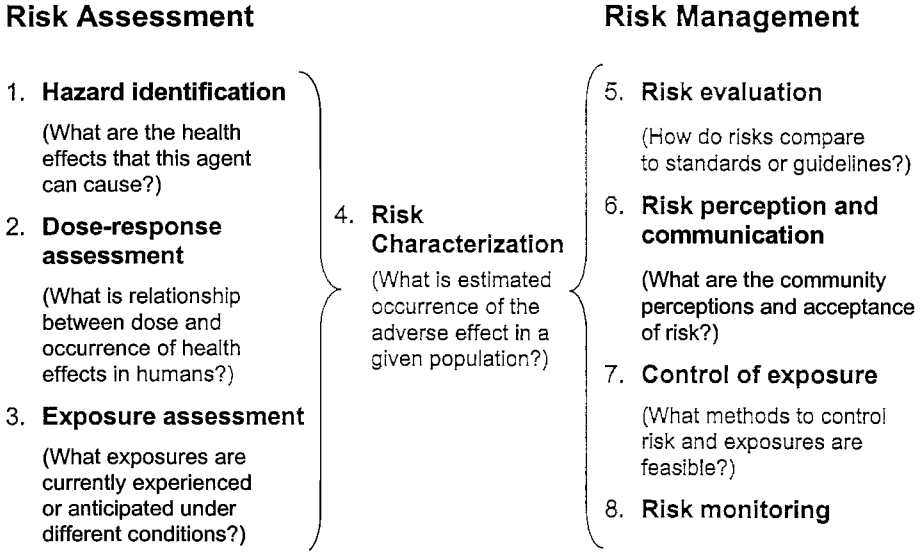
- the workplace environment: a person may spend a large part of his or her life in an occupational environment such as a coal mine or steel works which may be the site of

specific environmental problems and types of exposure; time spent in schools or other educational establishments can also be considered under this heading;

- the community (local) environment: a person may be exposed to air pollution caused by traffic, local industry or home heating, to chemicals leaking from a local waste dump, or to traffic noise — in the immediate area in which he or she lives;

- the broader environment: this can refer to the environment on a regional, national, international or global scale; associated exposures include photochemical air pollution, water pollution to rivers or the sea, radiation pollution from nuclear accidents, exposure to ultraviolet radiation as a result of stratospheric ozone depletion, and changes in vector-borne diseases due to global climate change.

Figure 1.1. Steps in Environmental Health Risk Assessment and Management



Source: adapted from Yassi, 1998.

Often the term general environment is used to identify all non-workplace environments (home, community and broader), and the term ambient environment has the same meaning. Commonly, an environment can be delineated by geographic boundaries. Thus environmental management activities may focus on individuals dispersed over a wide area, on groups living or working together, or on populations in defined areas or countries.

The health determinants that are not included in environment as defined above are: genetic factors, life-style and behavioural factors (involving clear individual choices e.g. smoking or drinking alcohol) and iatrogenic factors (exposures caused by medical interventions). The distinctions are somewhat arbitrary. For example exposure to cigarette smoke from smoking is considered a life-style factor, while exposure to cigarette smoke from someone else smoking is considered an environmental factor (and is usually referred to as environmental tobacco smoke).

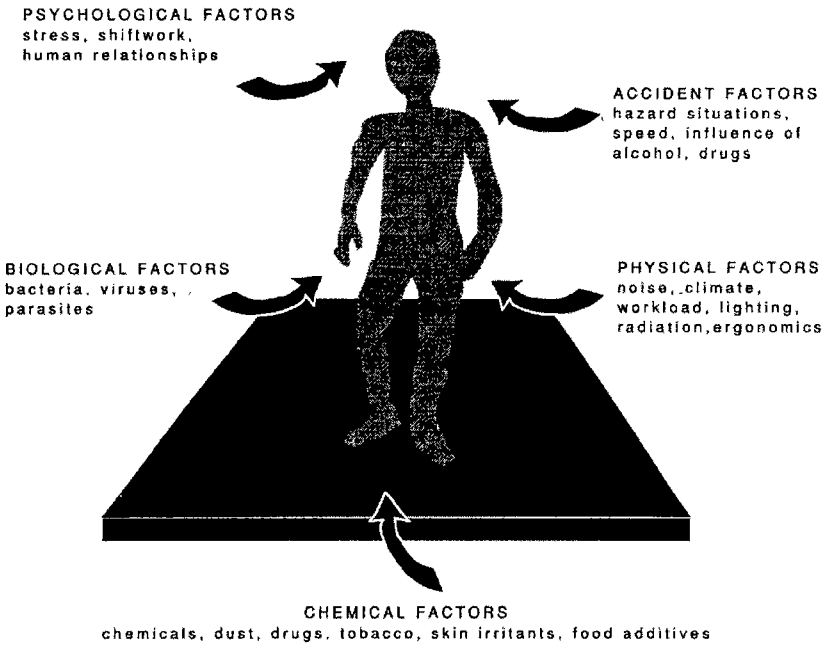
Epidemiology as a link between hazards, health effects and prevention

Epidemiology is one of the most important tools used in environmental health management owing to its capacity to assess environmental hazards as the potential cause of disease, and quantify their health impact on the population at risk. Thus it provides the information necessary for setting priorities for preventive action. It can also contribute to the monitoring and evaluation of the effectiveness of such action. In short, epidemiology is crucial to resolving environmental health problems.

Epidemiology is defined as: "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" (Last, 1995). Epidemiology therefore comprises not only studies, but also assessment of how study results have been used.

In environmental epidemiology the focus is on environmental hazards as determinants. We can classify them in a number of ways depending on their character or source — biological, chemical, physical, psychosocial, or safety-related (see Figure 1.2). The methods presented in this text can be applied to all of them, although since each country or locality has its own particular combination of environmental hazards, specific study designs for linking hazards, effects and prevention must be tailored to local conditions. As the focus here is on epidemiological investigation of the health impacts of "modern" types of environmental pollution, many of the examples used refer to chemical hazards.

Figure 1.2. Environmental factors that may affect health



Source: Beaglehole et al., 1993.

The specific methods for assessing the environmental health status of a locality and minimizing health risks vary. However, the core epidemiological approaches for studying and analysing the health risks are similar. For instance, if an outbreak of diarrhoeal disease occurs in a large school, the most likely cause will be contamination of drinking-water and/or food consumed at the school. Possible routes of transmission could be identified by comparing the children affected and the children not affected, with respect to what they have eaten and what they have drunk (a case–control approach). Suspected foodstuffs or drinking-water could be analysed on the basis of existing knowledge of the type of hazards that cause diarrhoea. Until this point, the approach is the same whether a biological factor, such as salmonella in food, or a chemical factor, such as copper or cadmium in drinking-water, is the suspected cause.

Further investigation might include analysis of the suspected drinking-water or food, using specific methods for each hazard; the epidemiological analysis of the results would be similar. However, the fact that biological hazards can spread from one person to another via infection would influence any further steps taken in an epidemiological investigation. If family members not attending the school develop diarrhoea a few days after the original outbreak, a biological hazard would be the likely cause. (If the poisoning was chemically-based, such secondary cases would not occur.) Preventive actions in the event of a biological hazard would include hygiene measures to protect family members from secondary infection. The source of bacteria in the kitchen area of the school would therefore require investigation and the methods used for cooking and storing foods would have to be reviewed. In the case of a chemical contaminant, the original contamination would probably have occurred during production of raw materials outside the school or during storage in containers. Follow-up action would therefore vary depending on whether it sought to eliminate infection caused by bacteria or poisoning caused by chemical contaminants. Thus epidemiology, by demonstrating whether the disease pattern is consistent with a fixed source of contamination or with spreading from person to person, establishes a linkage between hazard and effect that is essential when devising preventive actions.

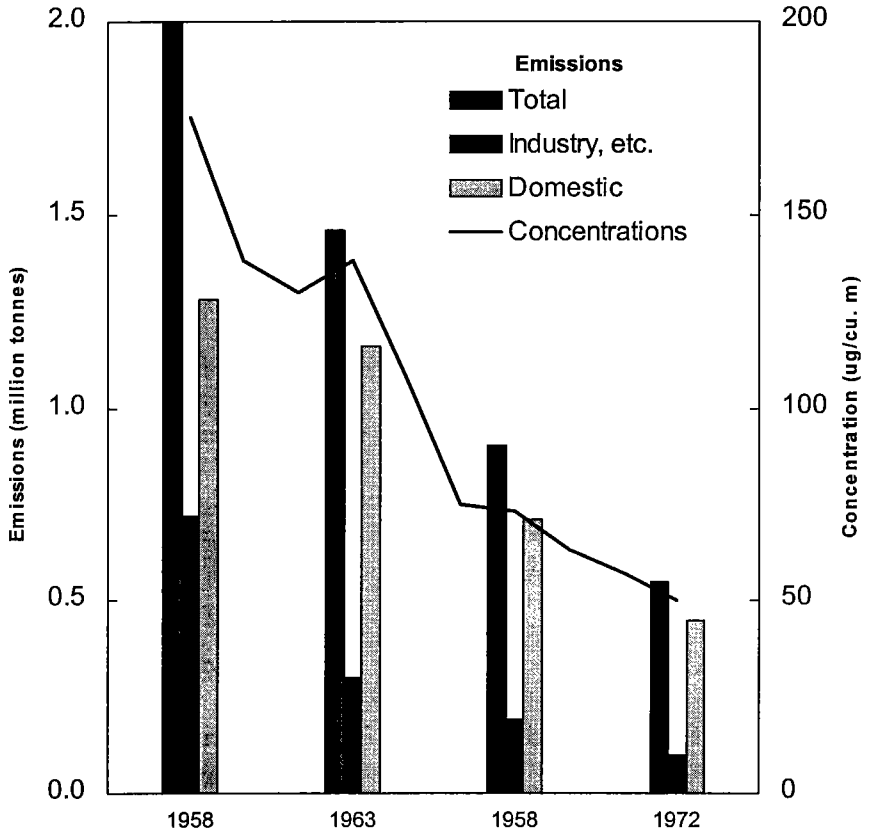
The establishment of a link between severe air pollution exposure in London in 1952 and increased mortality due to lung and heart disease is another example of the application of epidemiology to environmental health management (United Kingdom Ministry of Health, 1954). Epidemiological analysis of this pollution event led to continuous reduction of dust and sulphur dioxide pollution caused by coal burning through the establishment of "smoke-free zones", and progressive introduction of air pollution control equipment in industrial plants and power stations (Figure 1.3). These actions would not have been taken had the original study not demonstrated the link between air pollution and disease. Recently, re-analysis of mortality and air pollution data from the United Kingdom and the USA has shown that mortality is affected at low levels of dust and sulphur dioxide air pollution (Schwartz & Marcus, 1990). These epidemiological analyses have triggered reviews of air quality standards and led to calls for stricter standards.

Environmental epidemiology

As indicated in Box 1.2, environmental epidemiology is the branch of epidemiology that specifically deals with environmental exposures. A first set of guidelines on environmental epidemiology was produced by WHO in 1983 (WHO, 1983). Often, the home, community and general environment as well as the workplace environment are all included, but some scientists

separate occupational epidemiology (workplace environment) from environmental epidemiology. In this book we focus on the non-occupational environment. Environmental epidemiology can be defined as: *the study of the distribution of health-related states or events in specified populations in relation to determinants/hazards in the living environment of these populations, and the application of this study to the control of such hazards.*

Figure 1.3. Reduction of Sulfur Dioxide Air Pollution from Fuel Burning in the United Kingdom, 1958 to 1972



Source: Weatherley, 1976

Environmental epidemiology necessitates application of the ideas and methods of many scientific disciplines, including chemistry, meteorology, microbiology, and physics (to establish human exposure levels). It also uses elements of clinical medicine, biochemistry, and physiology (to establish the health impacts). By applying statistical and mathematical methods, environmental epidemiology integrates information collected via the other sciences. The interpretation of this information for preventive action is one of the most challenging and important aspects of environmental epidemiology.

It should be pointed out that in all epidemiology studies a variety of factors that influence health may need to be taken into account as each of them may contribute independently to the causation of the disease of interest. Environmental epidemiology studies therefore cannot ignore exposure to life-style factors such as tobacco smoking, dietary factors or underlying diseases or conditions

unrelated or not directly related to the environment. The factors that are not being studied can be dealt with as "confounders". Indeed, not infrequently, much of the effort in an environmental epidemiology study is directed towards measurement of non-environmental factors.

Box 1.2: Characteristics of environmental epidemiology

Environmental epidemiology:

- is concerned with the adverse health effects of exposures to environmental factors (primarily biological, physical, chemical)
- integrates information on exposure and health effects as well as other information from a variety of disciplines (such as toxicology, chemistry, ecology, audiology, radiation biology, clinical medicine, sanitary engineering, environmental engineering, industrial hygiene, meteorology, physical and social geography, acoustics, statistics and demography)
- is not only a set of methods for carrying out epidemiological studies, but a "mental strategy" for approaching the task of protecting people from environmental hazards
- represents a link between knowledge and action to protect health.

1.2 History and development of environmental epidemiology

Environmental epidemiology first emerged when Greek and Roman physicians and philosophers perceived links between features of the immediate environment and ill health. Nevertheless, it was not until much later that these links were evaluated quantitatively. Early examples of this quantitative approach include lists compiled of people, living or working in particular environments, who had died from a specific disease (e.g. by Bernardino Ramazzini in the 17th century who wrote the first text describing occupational diseases (Ramazzini, 1964, reprinted). The first scientific report identifying an environmental cause of cancer was made by Sir Percival Pott who described cancer of the scrotum in English chimney sweeps (Pott, 1775).

The first epidemiology study is also considered to be the first environmental epidemiological study, namely John Snow's study of cholera in London and its relationship to water supply location for different households (Snow, 1855). The health effect was an "infectious disease", which spread because of poor sanitation and drainage of sewage into the water supply. The epidemiological analysis included identification of where the affected people lived. As drinking-water is a typical environmental exposure route, Snow was able to demonstrate clearly the association between the source of drinking-water and the occurrence of cholera. This was a truly innovative analysis given that the existence of cholera bacteria had not yet been discovered. John Snow's study also shows that at this early stage of epidemiological development, the focus was on preventive action. John Snow did not know why water caused the disease. But clearly, initial preventive actions can often be taken even though complete information about the agent that is causing disease is not yet available. In this particular case, John Snow managed to stop the use of the contaminated water sources. This action reduced morbidity and mortality due to infectious disease long before the specific cause had been identified, and before antibiotics and vaccines had been developed.

Around the beginning of the 20th century, a series of developments in general medical understanding made the identification of specific causes of environmental disease easier. Koch, for instance, identified the role of bacteria in the late 19th century; his postulates for disease causation are still used when the cause of an outbreak of disease is being investigated. Then in the early 20th century, asbestos, lead, mercury and other chemical hazards were shown to be linked to specific diseases. The analysis of the data was crude and the conclusions were often drawn from studies of small groups of cases. But the notion of making comparisons between an exposed and a non-exposed population was becoming fairly common.

After the Second World War a number of well-publicized disease problems and studies brought a new urgency to the field of environmental health, and environmental and epidemiological monitoring became established tools. Examples can be found for all types of environmental hazards, be they physical, chemical, biological, psycho-social, or safety-related. For example:

- After the dropping of the nuclear bombs at Hiroshima and Nagasaki, a research institute dealing with epidemiological studies of radiation effects and treatment of the victims was established, the health of exposed populations was continuously monitored, and many research projects were carried out (e.g. Ishimaru & Ishimaru, 1975; Shimizu et al., 1990).

- In 1952, severe air pollution in London led to a dramatic epidemic of deaths from heart and lung disease. A detailed epidemiological study was carried out (United Kingdom Ministry of Health, 1954). It established a quantitative relationship between severe air pollution and the health effects that had been observed. Continued monitoring of air pollution and health in London and other UK cities proved the need for air pollution control; legislation to control air pollution in the UK was later enacted.

- In the 1950s, epidemiological and toxicological studies attributed two epidemics of unusual diseases in Japan to severe industrial pollution. One of these was the infamous "Minamata disease", caused by consumption of fish contaminated with methyl mercury (WHO, 1976 and 1990). (Minamata is the name of the city where the disease first occurred.) The other was "Itai-Itai disease", caused by consumption of rice contaminated with cadmium, in combination with poor nutrition (WHO, 1992a). ("Itai-itai", meaning "ouch-ouch", was the term used by patients to describe the severe pain they suffered as a result of multiple bone fractures.) The epidemiological studies showed that loss of calcium during pregnancy, and a lower calcium intake than that of men, rendered women particularly vulnerable to Itai Itai disease. Monitoring of the environment and health in the affected areas continues to this day.

- Legionnaires' disease is an example of an infectious disease, the cause of which was established by epidemiology studies. The disease was identified through an investigation of a pneumonia epidemic that affected more than 200 people and caused 34 deaths during an American Legion Convention at a hotel located in the United States in 1976 (Fraser et al., 1977). The causative agent was discovered to be a new species of bacterium, which was subsequently designated *Legionella pneumonphila* after this epidemic (Fraser, 1979). The discovery of *L. pneumonphila* led to the isolation of related organisms which is now recognized as a family of bacteria consisting of more than 50 species. *Legionella* infections account for up to 7 percent of community-acquired pneumonias and as much as 30 percent during hospital outbreaks.

1.3 Basic principles of environmental health

The basic premise of any environmental epidemiology investigation is that exposure to an environmental factor or factors can lead to health effects. The two concepts of **exposure** and **effect** are therefore integral to any environmental epidemiology study. Exposure is used to signify the contact that occurs between the human body and the environmental hazard of interest. Exposure can occur via: inhalation; ingestion; direct skin contact (or eye contact); transplacental blood contact (from mother to fetus), or injection directly into the body (although this may not be considered an environmental exposure in a strict sense). The amount of the hazard that has entered the body is termed the **dose**; the dose determines whether or not the individual exposed experiences an effect. Genetic variation in the population influences variation in internal dose and "susceptibility" to an effect. This genetic variation would appear as variability in the distribution of the dose in a typical sub-set of the population even if members of the population had similar exposures. (Chapter 2 discusses more about principles of exposure and Chapter 4 describes exposure measurement techniques for epidemiological studies.)

In this book the term **effect** is used as a generic description of any change in health status or body function that can be shown to be due to exposure to an environmental hazard. This term can apply to an individual as well as to a change in the average experience of a population. Some epidemiologists use the term only in relation to populations. We use the term **effect measure** to refer to those epidemiological variables that describe changes in population experience (see also Chapter 3).

Local effects occur at the site of contact with the environmental hazard. Inhalation of chlorine gas, for instance, causes severe damage to the tissues of the respiratory tract, which is the first site of bodily contact following inhalation. Irritation of the eyes after exposure to high levels of ambient ozone is another example of a local effect. Effects may also occur at sites other than the point of entry. A chemical, for instance, is transported (usually via the blood) after absorption, to other tissue. A **systemic effect** may then occur in this tissue.

Health effects can also be classified according to the speed with which they develop in the individual in relation to the beginning and end of the period during which exposure occurs: an **acute effect** is an effect that develops very rapidly after exposure starts whereas a **chronic effect** takes substantial time to develop. These effects can be **temporary** (or **reversible**), meaning that they disappear after exposure ends, or **permanent** (or **irreversible**), meaning that they persist after exposure has ended. (Chapter 2 discusses more about health effects and Chapter 5 describes health effects measurement strategies for epidemiological studies.)

These fundamental principles provide the foundation for interpreting links between the environmental conditions of a community and the occurrence of specific health problems. The environmental conditions of concern may be chemical factors, physical factors (such as ionizing radiation or noise), biological factors present in inert "vehicles" (such as bacteria or viruses in drinking-water and food), biological factors spread by live "vectors" (such as malaria plasmodia in mosquitos) and ergonomic safety factors (the conditions that lead to physical injuries). Many examples presented in this text are based on chemical exposures, for which the principles are most fitting. Nevertheless, when designing environmental epidemiology studies on other types of hazards, we can apply the same principles, with some adaptation.

Biological factors differ from the other types of factor in that they can multiply inside the human body and in "nutritious" materials (e.g. foodstuffs). This means that the presence of just one bacteria or parasite may lead to the development of the disease. For instance, a very low exposure to cholera bacteria may not in itself be sufficient to cause the disease, but eventually the individual concerned will develop the disease since the bacteria will multiply in the intestines. Another feature of biological factors is that they can spread from one human being to another and lead to "secondary cases". In addition, the exposure to most biological factors leads to the development of antibodies in the blood. An environmental epidemiology study of biological hazards may therefore focus on the identification of anti-bodies rather than identification of the hazard itself.

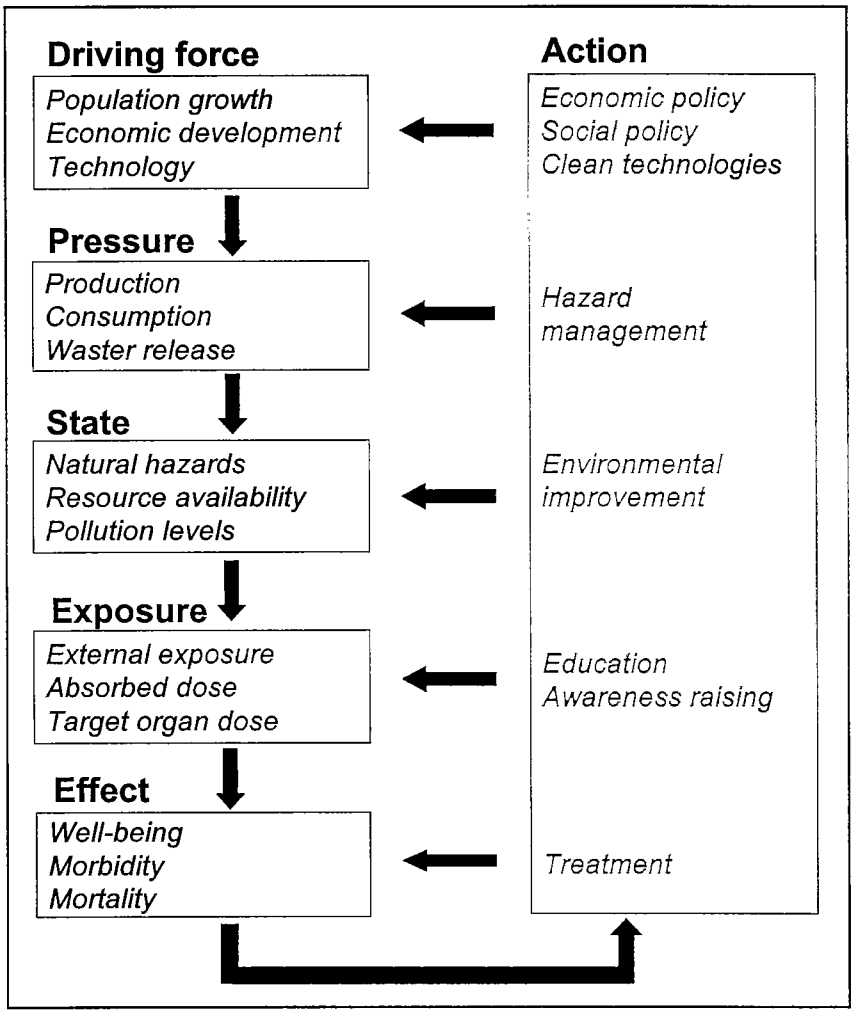
The different settings within which exposure occurs and the different routes of exposure constitute another of the important focal points of environmental health management. Most people do not spend all their time at the same site all day, but spend some time outdoors, some indoors, some at work, some at home, and so on. The exposure that occurs in each setting can contribute significantly to the **total human exposure**. Environmental epidemiology studies must take this into account. For instance, exposure to air pollution indoors in many Chinese cities is even higher than the exposure to air pollution outdoors (Chen et al., 1990). The total air pollution exposure is what determines the health effects, so both indoor and outdoor pollution need to be measured. The workplace is of particular interest since certain hazards occur there at much higher levels than in the general environment.

Measuring actual exposure is not always possible and factors leading to exposure may have to be used as **"surrogates" of exposure**. In the context of environmental health management these factors have been described as "driving forces" leading to "pressures on the environment", which in turn lead to changes in the "state" of the environment. These in turn alter the state of the environment by increasing existing exposures or introducing new ones, which produces a health effect. In order to rectify the problem, actions (i.e., environmental health management) must be undertaken at each level (see Figure 1.4). Thus, the framework is the Driving-force–Pressures–State–Exposure–Effects–Action (DPSEEA) (Kjellström & Corvalán, 1995). A challenge in environmental health is to identify the causal linkage between these forces or pressures and the actual exposures because this understanding will guide strategies to manage the environmental health problem. Another challenge for epidemiologists is to decide when some surrogates of exposure, such as the location of a person's residence, is a sufficiently precise indicator of environmental exposure that the surrogate can be used as a measure of exposure in an epidemiological study. The use of surrogate measures and other issues of measuring exposures are discussed in Chapter 4.

The association between environmental exposures and health effects can be complex because single environmental factors may cause **multiple effects**. For instance, lead causes biochemical alterations to the haem system eventually leading to anaemia, while at the same time affecting the central and peripheral nervous system. Such combined health effects can also be due to **combined exposures** to several hazards. In fact, in many cases of air and water pollution combined exposures do occur.

Very complex situations may arise if the exposure consists of a mixture of many chemicals, as at some hazardous waste sites. Studying the related health effects can prove extremely difficult. Combined exposures to biological hazards also occur, particularly in poor slum areas with poor sanitation and hygiene conditions.

Figure 1.4. **Development of environmental health indicators**



Source: Briggs et al., 1996.

The basic principles of environmental health will be referred to in the context of the different aspects of environmental epidemiology described in the other chapters of this book. However, the student of environmental epidemiology will need more detailed knowledge about environmental health science to make optimum use of this book. A recent textbook produced by WHO titled "Basic Environment Health" does provide the additional detail required (Yassi et al., 1998).

1.4 Environmental epidemiology and related sciences

Distinction between environmental epidemiology and occupational epidemiology

If environment is defined broadly, the workplace becomes one of the settings in which environmental hazards can occur. In many situations the borderline between workplace and general environment is unclear. Many people in developing countries are subsistence farmers and spend considerable time working in the fields surrounding their dwelling. Is this environment their workplace or their residence? Many women work in the home carrying out household tasks and taking care of children. Their home is also their work environment. This text presents methodologies that can be applied in any setting, but our examples focus on applications in the general environment.

The distinction between "environmental epidemiology" and "occupational epidemiology" is not always evident, and some people accordingly consider that occupational epidemiology is a component of environmental epidemiology. Yet occupational health has its own legal basis and considerable legislation exists relating to the protection of employees from harmful exposures. It also has a long tradition of specialized training and professional practice (for example, occupational hygiene, occupational medicine).

Application of epidemiology in the workplace differs from application of epidemiology in the general environment since the former focuses on people actively involved in "work". It can be expected, for instance, that neither children nor the elderly experience workplace exposures (although child labour, is still common in some societies). In addition, people suffering from disease or a handicap are often excluded from work. In short, the active workforce is generally healthier than the general population. Occupational epidemiology studies should therefore take into account the so-called "healthy worker effect" (McMichael, 1976).

The "healthy worker effect" is also dependent upon the relative quality of the environment at home and at work. In the early stages of industrialization (and particularly among "cottage industries") the general hygienic conditions of workplaces are usually paid very little attention by workplace owners. Crowding, heat, poor ventilation, unsafe buildings and poor access to drinking-water, are common environmental problems. As workplace standards improve these problems can be almost totally eliminated and, in some situations, the environmental conditions at work may even be better than in the workers' homes.

The workplace environment is often confined within buildings. As a result, exposure levels may be higher in the workplace environment than in the outdoor environment. Workplace exposures, be these inside or outside, may also be high if the source of the hazard is close to the worker (as in the case of farmers spraying pesticides). For many hazards that occur in air, the first indications of health effects in exposed populations have derived from occupational epidemiology studies; and since hazardous materials used in agriculture or industry are handled directly by workers, they often constitute the first population to be exposed. We can therefore regard the workplace as an indicator (or "sentinel") of what may happen in the general environment. Table 1.1 presents the steps in the development of our knowledge about cadmium poisoning, demonstrating how knowledge pertaining to cadmium poisoning in the workplace was acquired before knowledge pertaining to cadmium poisoning in the general environment.

Table 1.1. Development of knowledge of cadmium poisoning

Time period	Knowledge development
1930s	First **animal studies** showing lung effects
1940s	First studies of **workers** showing bone, lung and kidney effects (in France and Sweden)
1950s	First studies of **general population** (farmers) in Japan with kidney and bone effects (Itai-Itai disease) from exposure to contaminated water and rice
1960s	Detailed epidemiological studies of **general population** in Toyama, Japan, of Itai-Itai disease. Japanese government acknowledges link to cadmium exposure. Mechanism of kidney damage in **workers** understood.
1970s	First comprehensive review report published ("Cadmium in the Environment" by Friberg et al.) Epidemiological studies in several polluted areas of Japan; **general population.** First studies of **workers** linking cadmium exposure to lung cancer. WHO guidance on cadmium in food.
1980s	WHO guidance on cadmium in workplace environment. Epidemiological studies in polluted areas of China; **general population.** Mechanism behind links between cadmium-induced kidney effects and bone effects identified from **animal studies.**
1990s	WHO Environmental Health Criteria. International Agency for research on Cancer declares cadmium a human carcinogen.

Source: adapted from Friberg et al., 1986, and Kjellström, 1986.

From an epidemiological point of view, workers in a particular workplace form a well-defined and readily accessible group. Occupational epidemiology studies have therefore often been the first studies to be undertaken of the health effects of a particular hazard. Some occupational health investigations are in fact required by law. For instance, most countries have passed legislation concerning periodic screening of lead workers for lead exposures and health effects. Screening of workers for health effects is also common in other hazardous industries such as electroplating, mining and quarrying. The data from these screening programs and associated exposure monitoring can be used for occupational epidemiology studies.

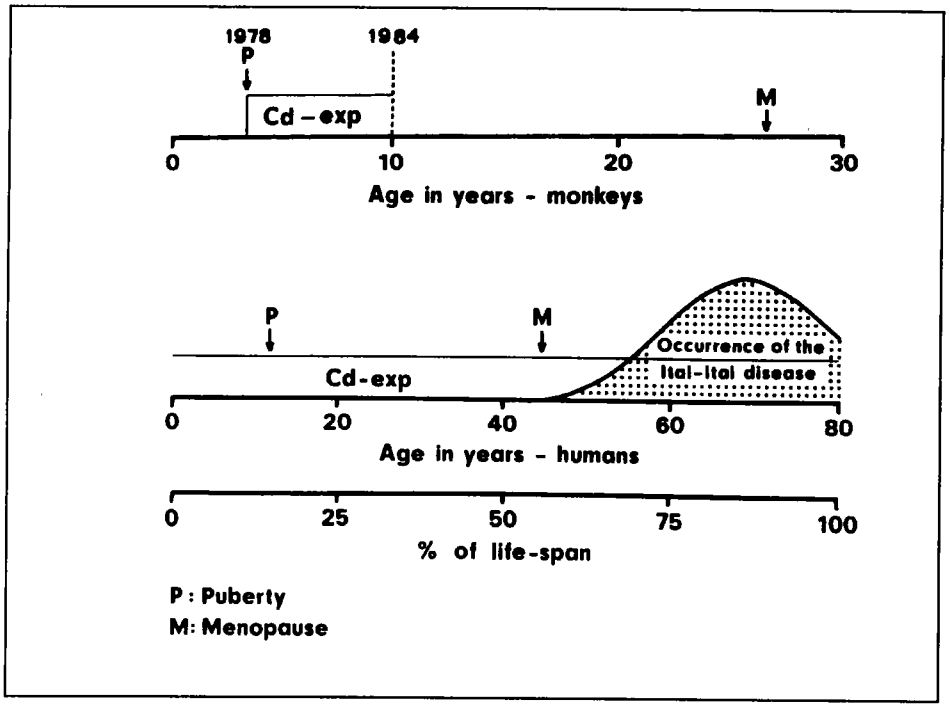
Methods used in occupational epidemiology have been described in a WHO publication (WHO, 1986) and numerous other texts.

Interrelationship between environmental epidemiology and experimental studies of animals

As described above, **exposure** and **health effect** are two of the fundamental concepts used in environmental epidemiology. They also form the basis of experimental studies, in which a test animal is exposed to an environmental factor until it experiences an effect. The experimental evidence is often used to assess the potential health risks to humans arising from such exposures. **Risk assessment** therefore involves combining toxicological and epidemiological information to make the most accurate assessment possible of human health risks.

Animal studies may indicate the human target organ in which health effects will occur and can provide useful qualitative information about the different steps in the metabolism of a hazard. If the animal has an organ structure similar to that of humans, it may well be assumed that the equivalent human organ would be affected in much the same way. However, the metabolism, size and life span of test animals differ from those of humans. An effect that is seen in humans after 30 years exposure (1/3 of a life-span) may require the same life-span fraction of exposure duration in the animal to demonstrate effects. In case of small animals, such as mice and rats, this period would be a year or less and is feasible to carry out. However, in larger or more long-lived animals, such as monkeys, this could mean decades of exposure in animal studies. Truly reflecting human exposure conditions may be very difficult to carry out. This is the case for the study of bone disease effects of cadmium exposure (Kjellström, 1986). (See Figure 1.5).

Figure 1.5. Comparison of life-span of humans and Rhesus monkeys



Source: Kjellström, 1986.

So major limitations exist with respect to incorporating quantitative conclusions, based solely on animal experimental data, in risk assessments. Furthermore, experimental studies often assume that the animals exposed are genetically identical. If they are indeed identical, use of small animal groups in studies is indicated. But if they are not, the end result could simply be that studies are not planned so as to take sufficient account of the variation in individual susceptibility in humans that is due to genetic factors. Moreover, in human populations, exposure may vary considerably among individuals. Epidemiological studies explore both exposure variation and variations in effects, and therefore represent real life more accurately than do animal experiments. If descriptions of the mechanisms of health effects are based on the results of both animal and epidemiological studies, they will be more useful for decision-making than those based solely on animal studies.

Experimental animal studies are also limited in that they focus on the specific effects of exposure to a single hazard. But in fact, exposures seldom occur in isolation. The effects of combined exposures must therefore be considered. Since epidemiological studies deal with real-life exposure situations, they consider the complexities of combined exposures. They can also provide evidence of the attributable risk and the actual number of people affected by a hazard (or combination of hazards) in a specific environment.

1.5 Applying environmental epidemiology principles in your work

Setting priorities for how and when to apply resources for planning and implementing environmental epidemiology studies is important. The first priority should be to help solve environmental health problems in the community. Studies might deal with a local situation, for which more information is needed in order to initiate effective preventive action, or seek information on how a particular hazard causes disease. Whatever the situation, though, the methods and guidance provided in this text will help the investigator collect information in a scientifically valid manner. The new information can then be interpreted and used to make decisions about actions to be taken to reduce or eliminate the disease risk.

In some situations, a simple descriptive study of the local situation can provide enough information for making the necessary preventive decisions. In others, a more detailed and complicated study may be required so that decision-makers are fully informed and pursue the most appropriate environmental health policies. The environmental epidemiologist must therefore know how to select the study design that best fits the situation.

Whichever study approach is taken, a report must be prepared that outlines the methods, results and interpretation of results. The study cannot be considered to have been completed until the community itself has been informed of the results. In addition, publishing the results in a scientific journal provides a permanent record of the findings, from which other investigators can learn.

Your purpose in studying environmental epidemiology may be to learn how to produce new data or new analyses for future environmental epidemiology studies, or to learn how to use data from such studies for decision-making in environmental health management. Alternatively, you may wish to learn more about this science so as to be able to teach it to students or junior staff in your institution. We can thus identify three major applications of environmental epidemiology:

namely, **research**, **practice** and **teaching**. A thorough understanding of the principles that underlie environmental epidemiology studies is crucial for each.

This text is designed to accompany a course that teaches students how to carry out an environmental epidemiology study and how to interpret studies. A number of additional learning materials could also be used in such a course. References to other relevant materials are included in this text. We make a point of referring to materials produced by WHO, since they are available in libraries throughout the world. (WHO provides many of its publications free of charge to many libraries in developing countries). A listing of further information resources and teaching materials is included in Chapter 12.

To fully understand this text it is necessary for the reader to have some background knowledge. First of all, knowledge of the basic principles of epidemiology is required, even though some of the key elements of this science are described in chapter 3. We suggest that the reader should already be familiar with the concepts described in Basic Epidemiology (Beaglehole et al., 1993). This knowledge required would be equivalent to a 20-30 hour course in epidemiology. Basic statistical methods also need to be understood to develop and implement the type of studies described in this text. In addition, it is important that the reader has a broad understanding of environmental health science principles and concepts. These can be found in the recent WHO document called Basic Environmental Health (Yassi et al., 1998). Again, this material would be equivalent to a 20-30 hour course.

Learning how to carry out environmental epidemiology studies is an ongoing process, however, which should include experience in planning and implementing studies, and in the reporting and interpretation of environmental epidemiology data. This book should therefore be seen as a tool to initiate development of deeper knowledge and understanding. Interaction with other scientists and public health staff is another important element in such a learning process. This text will provide a common ground in terms of concepts and terminology, so that communication among scientists about environmental epidemiology data is facilitated.

Chapter 2

Context of Environmental Epidemiology

Learning objectives:

- Understand how environmental health problems can arise in a community and lead to requests for an epidemiological investigation.
- Be able to describe the concepts of exposure, dose, effect and response, and their interrelationships, and to give examples pertaining to different environmental media.
- Understand the distinction between biomarkers of exposure and biomarkers of effect.
- Appreciate the social, economic, ethical, legal and scientific issues involved in conducting an environmental epidemiology investigation.
- Understand the process for deciding whether a study should be undertaken.

2.1 Context of environmental epidemiology studies

As described in chapter 1, environmental epidemiology deals with all types of environmental hazards, each of them requiring distinct approaches to measurement of exposures and effects. When the occurrence of effect is dramatic in nature (an "epidemic") and is caused by known hazards, the epidemiological methods may be quite simple, and information to guide prevention can be rapidly obtained. In many instances environmental epidemiology studies require more sophisticated techniques and are more difficult to carry out, for the following reasons.

Firstly, exposure to environmental contaminants may occur at the individual level, but more often occurs at the level of the household, community, or even region so it can be difficult to measure precisely an individual's exposure. Great attention must therefore be given to accurate exposure assessment and the possibility of multiple exposure routes and combined exposures to different hazards causing similar effects.

Secondly, the excess risks associated with exposure in the general population may be quite small. Yet even modestly elevated risks can represent a significant impact on the population if large groups are exposed. For example, recent research has suggested that small increases (10 $\mu g/m^3$) in the concentration of respirable particulates in the air may be associated with excess mortality of one to two percent (Dockery & Pope, 1994) which may represent a large number of excess deaths if entire urban populations are exposed. Detecting such small, but significant excess risks can be difficult though due to the possibility of random error and bias. Qualitative and quantitative assessment of the influence of bias is therefore crucial. Indeed, the need to improve the ability of epidemiological studies to detect small effects through, for example, enhanced validity and precision of exposure estimates and increased sensitivity and specificity of outcome measures, is great.

A third feature of environmental epidemiology is the interest that is often shown by the public and regulatory agencies in its findings. Documentation of an adverse health effect due to an environmental factor, such as industrial waste or air pollution, may have significant ramifications for public policy or financial liability. To ensure that decision-makers understand the results of the environmental epidemiology study, reports have to be clear and written in understandable language. This is the basis for good risk communication. The ability to interact effectively with all interested parties is also an essential element of many environmental epidemiology studies.

Impetus for study

Epidemiological studies of environmental health issues are undertaken for a variety of reasons. Often they aim to establish or improve scientific knowledge of the causal relationship between environmental exposure and disease. Such studies are often undertaken by a university or research institution. If successful, awareness of the health impacts of environmental hazards and the need to prevent or minimize them is increased. An epidemiological study might also focus on obtaining descriptive data for initial decision-making regarding preventive action. Such investigations may be conducted by local or regional public health agencies in response to community concerns about a perceived environmental exposure or disease epidemic. These investigations must have a practical, problem-solving orientation to address the community's concerns in the shortest possible time period. More detailed epidemiological investigation, to identify the precise causes of ill health and to monitor the interventions, could be pursued later.

Alternatively, the impetus for a study may derive from unanswered questions about the health impact of an environmental hazard that was discovered through previous research, or following field experiences of environmental diseases. Community awareness of environmental health issues, which has increased significantly in recent years, and led to a demand from politicians as well as communities for research findings that are demonstrably valid, can also create a need for epidemiological research. In particular instances this may necessitate additional studies that are more detailed than original studies that indicated or highlighted an environmental health problem. In general, the nature of an environmental epidemiological study is determined by whoever is responsible for conducting the study, the purpose of the study, and the resources available to the investigators.

Many diseases may be related to changes in the environment (Table 2.1). Thus, excess occurrence of these diseases would generally provide motivation for conducting an environmental epidemiology study. The causation may be multifactorial, with an influence by environmental, behavioural ('lifestyle') and socio-economic factors.

The exposure of a population to a particular factor of national or international interest may also prompt further environmental epidemiological study. In Mexico City, for instance, ozone concentrations can be very high and have received much national interest, making this city an appropriate place in which to conduct research on its health effects (Castillejos et al., 1992).

Table 2.1. **Disease categories that have been related to environmental agents**

Disease Group	Subgroup	Environmental factor
Gastrointestinal	Diarrheal disease	Bacteria and viruses in water and food
Cancer	Lung	Tobacco smoke, metals, radiation
	Blood cells	Hydrocarbons, radiation
	Soft tissue	Hydrocarbons, herbicides
	Liver	Hydrocarbons
	Urinary tract	Certain chemical compounds
Respiratory	Asthma	Climate, particulates, ozone
	Bronchitis	Particulate, sulfates
	Emphysema	Smoke, sulfates, ozone
Reproductive	Malformations	Chemical mixtures, solvents
	Abortions	Metals, hydrocarbons
	Low birth weight	Tobacco smoke, chemical mixtures
Neurological	Development	Lead
	Transmission	Lead, organic solvents

Epidemiological studies are also often carried out in response to community concern about newly developing or increasing incidence of health effects that are believed to be linked to changes in environmental conditions that have occurred following industrial or agricultural developments. Cases of a disease can occur as an "epidemic" — for example, a set of congenital malformations may be identified within a community hospital over a short period of time (Walss-Rodriguez et al., 1990). But if the excess risk is small, cases may accumulate gradually over an extended period. They may nevertheless represent a considerable social and economic burden for the community. Emphysema in industrial towns is one such example. Similarly, clusters of severe illnesses (e.g. cancer) that were not observed previously in a town often initiate investigation of environmental factors. In the USA, in particular, hazardous waste sites are currently causing much disquiet, particularly among local communities. Many of these sites are potential sources of exposure to mutagenic and carcinogenic agents, neurotoxic agents, and respiratory and ocular irritants. Some of them have undergone epidemiological scrutiny. In New York State, for example, a study revealed a 6% to 15% increased risk of congenital malformations in children born to mothers living close to such exposure sources (Geschwind et al., 1992).

In the event of an environmental disaster, the need for environmental epidemiology investigations may be particularly urgent (Kreimer & Munasigne, 1991). Similarly, cumulative exposure following an ecological disaster may require epidemiological investigation, as in Hungary when an unusual cluster of congenital abnormalities was suspected to be related to trichlorfon contamination of fish (Czeizel et al., 1993). A perceived disaster may also call for epidemiological research, as in the case of health effects observed following the release of chemicals from a pesticide and herbicide plant in Veracruz, Mexico (Restrepo, 1993).

Studies are sometimes undertaken in response to public concern even if voiced by a comparatively small group, such as those who claim to have been poisoned by mercury from dental fillings or affected by electromagnetic fields (EMF). Resultant preventive action, however, may protect the health of large numbers of people. In Sweden, about 70 new cases of child leukaemia are reported every year, and it has been estimated that about one extra case of leukaemia may occur each year as a result of EMF exposure (Feychting & Ahlbom, 1995; Feychting et al., 1995) . The causal nature of the association is still unclear, but the estimate provides an idea of the relative magnitude of this public health problem compared to other causes of death among children in Sweden. Nevertheless, epidemiological studies of electromagnetic radiation are being undertaken in response to the public concern about this exposure. Information obtained by these studies should be generalizable to other populations.

Environmental epidemiology studies can also contribute to practical environmental health management, which requires continuous monitoring of environmental exposures and health effects. Such monitoring is one means of collecting quantitative data that can be analysed epidemiologically and prevented so that they can be readily interpreted by decision-makers and communities. As discussed in Chapter 1, regulators also often require specific environmental epidemiological research in the form of human health data collection to support environmental management and policy decisions. Ideally, an epidemiological study is designed and implemented so that it meets both practical management needs and scientific needs. Admittedly, it can be difficult to combine the strict criteria of a scientific study with the flexibility and openness required for a management study involving a community.

Multi-sectoral interaction

Environmental epidemiology depends upon the collaboration of professionals from multiple disciplines and interaction between numerous sectors of society. Initially concerned primarily with sanitation problems, the focus of environmental epidemiology has broadened considerably, so that environmental epidemiologists must now work with professionals from sectors such as housing, farming and food supply, transport and industry. Assessment of the potential impact on human health of a factor such as industrial development can be very complex and may involve balancing the potential adverse effects of exposure to industrial contaminants against the positive health effects associated with economic development. A society may also need to make difficult decisions regarding the allocation of limited resources when addressing human health concerns and other needs. For example, a choice may have to be made between implementing restrictive environmental health policies regarding toxic contaminants and using resources for housing or education.

Given its broad focus, environmental epidemiology can contribute to many disciplines, to development initiatives such as Healthy Cities (see Box 2.1.) and to multidisciplinary concepts such as that of "Urban Metabolism" whereby the functions of a city are viewed in terms of its social, physical, economic, demographic and epidemiological interactions (Hancock, 1993; Ness, 1994). Ultimately, environmental epidemiology will expand its reach yet further leading to an increased understanding of the interactions between the multiple factors affecting health and well-being, and the capacity to propose interventions to alleviate difficult population conditions, such as those found in many of the world's large cities.

Box 2.1. The Healthy Cities movement

The Healthy Cities Projects promote collaboration between the key agencies (government, business, community organizations, professional groups and nongovernmental organizations) of a city so that they can agree on how they can jointly improve the quality of their city's living environment. The movement started as a European project to create action-oriented approaches to urban health. In recent years all WHO regional offices have developed healthy cities projects in collaboration with the cities and municipalities of a large number of countries.

The project's primary objectives are to strengthen the capacity of municipal governments to address health and environment problems, and to provide opportunities for individuals, families, and community groups to participate in resolving those problems. Healthy Cities projects provide a framework that combines several elements: increased awareness of health and environment issues relating to urban development; a network of cities to promote information exchange and technology transfer; and linkage of health and environment technical programmes with political mobilization and community participation. Environmental epidemiology studies and surveys contribute to the projects by providing information for decision-making.

Source: WHO, 1992d.

Social, economic, legal and policy aspects

In some countries, environmental health risks are receiving increasing attention in the mass media with the result that alerts regarding new threats to health from environmental factors appear frequently. Very often though the concern is based not on evidence from epidemiological studies, but on evidence from animal toxicology or from case reports of only a few possibly affected individuals. As a rule, the public health implications are not considered by the media in reporting these health risks. These implications need to be analysed in order to set priorities for prevention. Rational decision-making in relation to prevention would benefit from a greater reliance on environmental epidemiology. However, several factors complicate the use of environmental epidemiology as a basis for health risk assessment. These include, for example, the long latency period of some effects — epidemiological studies may not yield informative findings until many years after a hazard has been introduced into the environment. Moreover, epidemiological studies can be quite costly, meaning that evaluation of all the many potential environmental health risks faced by a population may not be feasible.

Environmental epidemiology has the capacity to provide information that can contribute to rational decision-making, although the use made of this capacity varies. Some societies are very concerned about environmental hazards and act accordingly. They instigate environmental epidemiology studies and use the results of these studies to establish guidelines and regulations regarding potentially hazardous exposures. Other societies fail or are unable to develop environmental epidemiology capacity, and consequently either remain "unaware" of environmental hazards, or do not act to prevent or control them. At the core of the discussion around the epidemiological transition (Bobadilla et al., 1993), environmental epidemiology has contributed to building consensus on risk affecting populations. Through this involvement, goals for safety standards, that are in accord with local values, can be set.

Environmental epidemiology can contribute to rationale allocation of resources by providing quantitative estimates of the risk reduction that could be anticipated by controlling exposures to environmental hazards. Risk estimates derived from epidemiological studies can, therefore, be used for cost-effectiveness analysis by environmental managers.

Regulators often require environmental epidemiological research in the form of human health data to inform policy decisions. Causal relationships are most convincingly established by human data (see e.g. IARC, 1987), even if experimental data are negative, as with arsenic carcinogenesis (IARC, 1980; Rossman, 1998). Projections based on animal experiments, or extrapolations from occupational epidemiology studies involving higher levels of exposure than occur among the general population are less reliable than direct epidemiological observations in the general population.

Epidemiological studies are also sometimes used to improve the quantitative basis for risk assessment and standard setting. The focus is then on exposures and effects that have already been identified. However, the original study groups may have been small and the range of exposure levels limited. Repeating a study, using the same methods as the original reported study, can produce very useful additional information. The repeat study can be carried out in the same population setting or in a different setting, perhaps in another country. For example, much epidemiological information on major environmental hazards, such as ambient sulphur dioxide, is based on studies conducted in the 1950s and 1960s in industrially developed countries. Conducting studies in developing countries where the exposure levels of these pollutants have become very high would confirm and expand this quantitative knowledge base, and provide further information concerning the extrapolation of risk estimates from one population to another.

In North America, evidence from epidemiological studies is sometimes required in legal proceedings. For the environmental epidemiologist, this may necessitate use of particular data management and documentation methods — ideally, these should be considered at the planning stage of studies.

Risk perception and communication

Environmental epidemiology takes place within the context of risk so the epidemiologist must learn how to assess risk and how to communicate risk (Baker, 1990). It is essential to differentiate between "hazard" and "risk." In this book, we use the definitions established by WHO (1989). Hazard is a source of danger; a qualitative term expressing the potential that an environmental agent can harm health. Risk is a quantitative probability that a health effect will occur after a specified "amount" of a hazard has exposed an individual. Risk assessment is the qualitative or quantitative estimation of the likelihood of adverse effects that may result from exposure to specified health hazards or from the absence of beneficial influences. The transfer of risk information from an expert to a non-expert audience is known as "risk communication".

Risk perception

A prerequisite of successful risk communication is an awareness of the extent to which risk perception is culturally dependant. In other words, risk perception is part of the value system of a society (Fischoff, 1993). For example, in very impoverished societies, typhoid fever might be

perceived as being of low risk given the existence of much more "immediate" risks to health such as malnutrition or inadequate shelter. But in highly industrialized societies, even a small risk of contracting typhoid fever would be of great concern. (Table 2.2 presents a number of factors that may influence perception of risk.) Risk perception is not a scientific construct. The contrasting and comparing of risks, and the setting of uniform levels of acceptable risk, are accordingly complex tasks.

Table 2.2. Factors related to the perception of risk

Catastrophic potential	Potential for voluntary control
Unfamiliar events	Willingness to be exposed
Understanding of event	Level of certainty
Latency	Effects on future generations
Equity	Years lost of healthy life
Trust in institutions	Media awareness
Perceived benefit of risk-taking	History of other catastrophes
Vulnerability	Natural origin
Quantifiability	Percentage of population exposed

Risk communication

Risk communication varies according to its level, type, purpose, methods, and the people who undertake it. Depending on the number of people at whom a communication is targeted, and the specificity of its content, a communication might be intended to inform the individual, particular organizations, or the general population. Other target groups include the family, the work group, or the classroom.

It is a mistake to consider any society as homogeneous when planning risk communication (Montoya, 1992). In most regions of the world societies are educationally heterogenous. Media risk communication programmes must cater not only for the illiterate, or for partly educated populations, but also for those who are well educated. Understanding the socio-cultural context and the target audience's position on the various aspects of an issue are also important. Techniques such as questionnaires, nominal groups, focus groups or general surveys, often prove helpful when assessing the information needs of a particular audience or target group. For instance, when discussing the manufacture of glazed pottery and the risk of lead contamination to which this process exposes them, Mexican craftsmen claimed that a change in technology would not be to their benefit, since it entailed no risk. They argued that lead intoxication was just a "physician's concern". Implementation of effective preventive measures only became possible after an environmental epidemiological study had demonstrated quantitatively to the craftsmen and their families the health consequences of lead exposure within the household (Santos-Burgoa, unpublished information).

Risk communication can be carried out orally, or via printed media (e.g. newspapers, booklets), visual media (e.g. posters, television advertisements), or even electronic media. Computerized interactive programmes, for instance, have been used to explain water risks to members of the

general population (Kubek & Robillard, 1993). Methods of oral communication include individual and group meetings, as well as interviews with media representatives.

Effective risk communication involves coherent messages from a variety of sources. The contributions that physicians, social workers, natural leaders and professional communicators can make to it are important (Sparks and Cooper, 1993). The environmental epidemiologist needs to provide the communicators with as accurate risk estimates as possible.

2.2 Biological basis for environmental epidemiology

Exposure and dose

The central role played by exposure was already referred to in Chapter 1. But before considering the methodological problems of exposure assessment, the relation of **exposure** to **dose** must be understood. These terms are often used interchangeably and incorrectly by both epidemiologists and other environmental health scientists. Exposure refers to the concentration of an agent in the environment that comes into contact with the external portion of the human body, while dose is the amount of the agent that actually enters the body. The concept of dose may be further defined as the **target organ dose**, which is the amount of the agent that reaches the susceptible organ or tissue within the body. The term "target organ dose" is thus used by various investigators to refer to the uptake of an agent by a specific organ, tissue or cell, or the concentration of that agent in the relevant organ, tissue or cell.

It is not until a pollutant reaches the internal portion of the body that the exposure can be expressed as a dose. A person could dive into polluted water and be highly exposed to the pollutants therein. Yet a lung dose would occur only if the persons aspirated the water.

In epidemiology, exposure estimates are usually based on data collected by environmental monitoring instruments or on estimates obtained by direct or indirect methods. These measurement techniques are discussed in Chapter 4. Monitoring data are used to estimate exposure, exposure is used to estimate dose, and dose is used to estimate the biologically effective dose at the target organ. Failure to understand the distinction between these terms can have an adverse effect on the design and interpretation of an environmental epidemiology study.

Health effects

Health effects of interest in environmental epidemiology cover a wide spectrum — from subjective annoyance to manifest disease or death — and may involve more than one organ system (for example, the respiratory tract and the central nervous system). (Measurement of health effects is discussed in Chapter 5.) In recent years, following developments in molecular biology and toxicology, interest has focused increasingly on measurable pathophysiological changes (subclinical signs) and early adverse effects, rather than on overt clinical disease. Examples of early adverse effects include biological indicators or "markers" for genotoxic effects, immunotoxicity and neurological damage. Markers of early effects can be a useful aid to prevention, particularly if they relate to reversible changes.

As mentioned in Chapter 1, health effects caused by environmental factors may be acute or chronic. For example, ambient air pollution can lead to acute lung function disturbances and to lung cancer. Chronic effects are often associated with long-term exposure, but at levels below those at which acute effects are observed. Repeated exposure may cause a reversible effect, but may also induce irreversible damage — for example, permanent hearing loss following repeated noise exposure. Studies of reversible health effects feature prominently in current environmental epidemiology research and have stimulated the development of special methods for study design and analysis. This emphasis is because findings of these studies may identify strategies for secondary prevention or treatment of the reversible effects.

Health effects induced by environmental factors may have a long latency. Follow-up of the atomic bomb survivors in Hiroshima and Nagasaki showed that latency periods for cancer, following exposure to ionizing radiation, ranged from a few years (leukaemia) to decades (solid tumours) (Miller,1984). Identification of early subclinical signs of toxicity is therefore important, particularly if late effects are likely to be severe. However, inadequate exposure data and the increased difficulty of controlling potential confounding variables render epidemiological studies of health effects with long latency particularly vulnerable to bias.

Health effects may be **local** (i.e. they occur at the site where the agent enters the body), or **systemic** (i.e. they affect organs inside the body after absorption of the agent). Understanding of systemic effects requires knowledge of factors relating to uptake, metabolism and excretion. The **critical organ** is the organ that first displays adverse effects following exposure to a harmful environmental agent. For example, the brain is the critical organ for methyl mercury and lead. Epidemiological studies of the effects on the critical organs of exposure to environmental factors are especially useful for establishing exposure guidelines and standards.

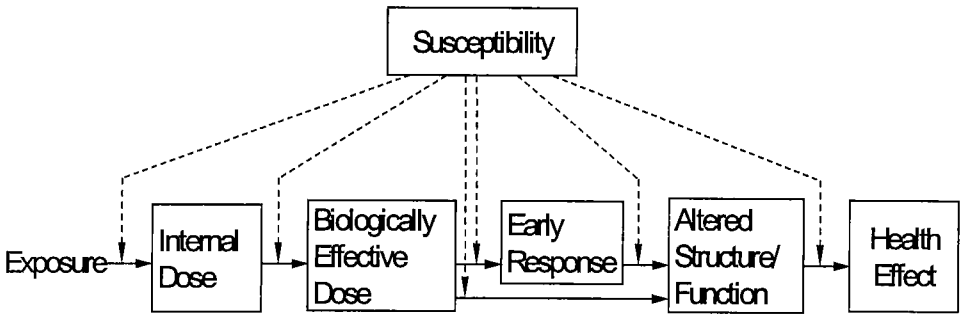
Biomarkers

A **biological marker** or biomarker may be defined as any measurable biochemical, physiological, cytological, morphological, or other biological parameter obtainable from human tissues, fluids, or expired gases, that is associated (directly or indirectly) with exposure to an environmental pollutant (Griffith et al., 1989). Many if not all tissues in the body are subject to adverse effects following exposure to environmental contaminants. The tissues and fluids used to measure biomarkers include, blood, urine, faeces, teeth, hair, saliva, amniotic fluids and cells, and semen. In addition, fingernails or toenails (to measure selenium, for instance) and subcutaneous adipose tissue (to measure fat-soluble organochlorine pesticides when biopsy specimens are available) can be used to measure exposure. Effective sampling sometimes necessitates use of invasive techniques. Blood and urine are the materials most commonly used as biomarkers. Descriptions of several kinds of biomarkers can be found in Hulka et al. (1990), Schulte (1987) and Schulte & Perera (1993). Figure 2.1 presents the relationship between exposure and dose, early biochemical/physiological changes, altered structure, and disease.

Biomarkers of exposure include measurements of pollutants or their metabolites in body tissues and fluids, and reflect the pollutant amounts that have entered the human body – thus, they should be considered indicators of dose, but "biomarkers of exposure" has become the generally used term. In principle, such biomarkers are more direct determinants of human health than concentrations of pollutants measured in the environment. In effect, biomarkers of exposure

integrate exposure from all environmental media and pathways. This integration can be a disadvantage, however, if the aim of an environmental epidemiology study is to evaluate the importance of individual environmental exposure pathways. In such cases, the relationship between the marker of dose and exposure via the separate environmental pathways may need to be determined (Brunekreef et al., 1983).

Figure 2.1. Spectrum from exposure to health effect



Source: adapted from NRC, 1987.

Table 2.3 illustrates the biomarkers used for studying exposure to environmental tobacco smoke (ETS) (i.e. tobacco smoke in air that may be inhaled by a person other than the person smoking), and gives the values for nicotine and cotinine in plasma, saliva and urine in smokers, and in two categories of nonsmoker (i.e. exposed or not exposed to ETS).

Table 2.3. Nicotine and cotinine as biological markers of exposure to tobacco smoke

	Nonsmokers without ETS exposure	Nonsmokers with ETS exposure	Active smokers
	Mean (No. = 46)	Mean (No. = 54)	Mean (No. = 94)
Nicotine (ng/mL)			
in plasma	1.0	0.8	14.8
in saliva	3.8	5.5	673
in urine	3.9	12.1*	1750
Cotinine (ng/mL)			
in plasma	0.8	2.0*	275
in saliva	0.7	2.5**	310
in urine	1.6	7.7**	1390

Differences between nonsmokers exposed to ETS compared with nonsmokers without exposure:
* $p<0.01$; ** $p<0.001$.

Source: adapted from NRC (1986); Jarvis et al. (1984).

Each type of biomarker has its own characteristic kinetics, determined by factors such as water solubility, excretion, accumulation in tissues and metabolic transformation. The extent to which a biomarker represents different durations and timing of exposure is of importance. Thus carbon monoxide (CO) in exhaled breath would represent exposure to CO that had occurred during the preceding few hours. So it would only be worthwhile measuring CO in exhaled breath if the study focusses on relatively acute effects. Conversely, bone or dentine lead would represent exposure to lead that had occurred over several years. This is because lead binds itself to these tissues and once it has entered the body is stored there for many years. These biomarkers of lead exposure can be used to study chronic effects (e.g. neurological and renal damage).

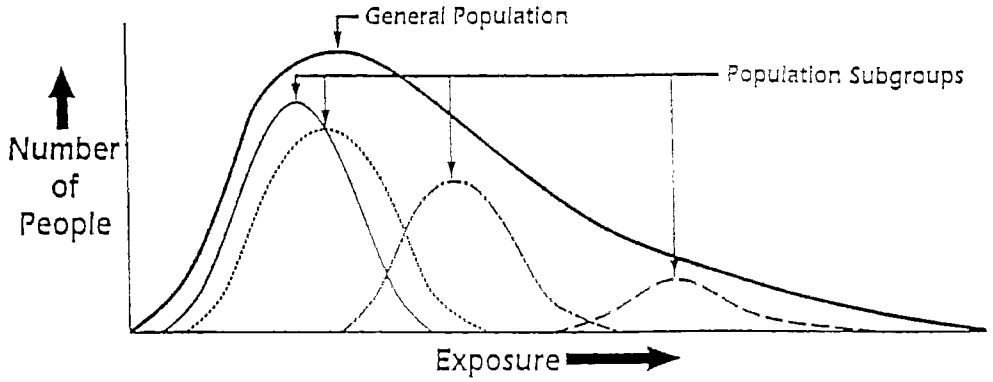
Box 2.2. Lead exposure and intellectual development of children

The Port Pirie Cohort Study examined the relationship between environmental exposure to lead and children's intellectual development. It is an example of an environmental epidemiology investigation based on repeated measurement of a biological marker of internal dose (blood lead). Lead was measured in the blood of the mothers before they gave birth, in umbilical cord blood, and in blood sampled from each child at the ages of 6, 15 and 24 months, and once every year thereafter. A detailed longitudinal record of internal dose was created, which was then linked to observations of the children's intellectual development.

Source: Baghurst et al., 1992.

Measuring "exposures" with biomarkers of internal dose makes it possible to acquire relatively accurate individual exposure levels. Within any population these individual levels always display a variation, often following normal or log-normal distributions (Figure 2.2). Variation among specific sub-groups may identify groups with different exposure levels.

Figure 2.2. Normal and log-normal distributions in populations



Biomarkers may also reflect early adverse effects or organ system damage. Markers of reproductive effects, for example, include reduced sperm count, changes in sperm morphology, mutagens in body fluids, and somatic cell mutation. Biomarkers based on immunological

parameters, such as antibodies, are used to evaluate infectious diseases and hypersensitivity (allergy) diseases. Biomarkers of neurological effects include depressed acetylcholinesterase in serum and plasma following exposure to an organophosphate pesticide. Several measures, such as DNA- or protein adducts, and sister chromatid exchange (SCE), have been used as biomarkers of carcinogenesis (Perera et al., 1987; Hulka et al., 1990; Perera et al., 1991).

Biomarkers of susceptibility are used to show whether a person has genetic or environmentally derived variations in metabolism of a chemical that could influence the impact on the target organ to environmental exposure. Genetic polymorphism in microsomal enzymes, such as cytochrome P_{450}, which influences the metabolism of many toxic chemicals is an example of a potential biomarker of susceptibility. People with certain genetic forms of these enzymes have a reduced ability to produce proteins or enzymes needed to metabolize a foreign chemical. In epidemiological studies, susceptibility biomarkers can improve the precision and strength of putative exposure–disease associations. This is because they avoid the dilution effect that occurs in populations that contain a large proportion of non-susceptible persons and clarify variation that may be due to differences in susceptibility rather than to differences in exposures.

Epidemiologically, a biomarker may serve as the dependent variable in one design application and as the independent variable in another. For example, in a study designed to observe the effects of phosphine (a chemical fumigant) exposure on the rate of sister chromatid exchange (SCE) in workers employed in grain elevators, the SCE biomarker could serve as the dependent (or outcome) variable. However, in a study designed to investigate the relationship of phosphine exposure to the occurrence of cancer, the biomarker (SCE) could effectively be used to indicate exposure and serve as the independent (or predictor) variable. Viewed as a dependent variable, a biomarker can serve to promote protective measures, up to and including modification or termination of hazardous exposure. As an independent variable, a biomarker may be viewed as the predictor of an effect and may promote understanding of the relative contribution of precursor exposures to the etiology of the disease or effect in question.

Irrespective, however, of whether a biomarker is to be used as a dependent variable or an independent variable, it should be judged in terms of its **validity** and **reliability**. (These terms reviewed in Chapter 3.) Validity incorporates the concepts of sensitivity and specificity. Reliability refers to reproducibility. If biomarkers are to be used in an environmental epidemiology study, "sensitivity" and "specificity", in particular, must be clearly understood. Sensitivity refers to the proportion of truly exposed people identified by the biomarker as exposed. Specificity refers to the proportion of truly unexposed people that are defined as unexposed by the biomarker. In the context of biomarkers these concepts are also applied to the laboratory methods. Laboratory sensitivity refers to the ability of the detection system to record the presence of the biomarker. Laboratory specificity refers to the ability of the detection system to fail to record the biomarker when it is absent.

Schulte (1987) has suggested that biomarkers should not be used in epidemiological studies until validation studies, including determinations of sensitivity, specificity and predictive value, have been completed. Validation studies should focus on the "normal population range" of the findings, adequate sample size and the control of confounders. Additionally, persistence of the biomarker in terms of recent or past exposure, as well as a plausible explanation for the biological mechanism involved in the suspect event should be determined through laboratory and experimental studies, prior to the inclusion of the biomarker in epidemiological studies.

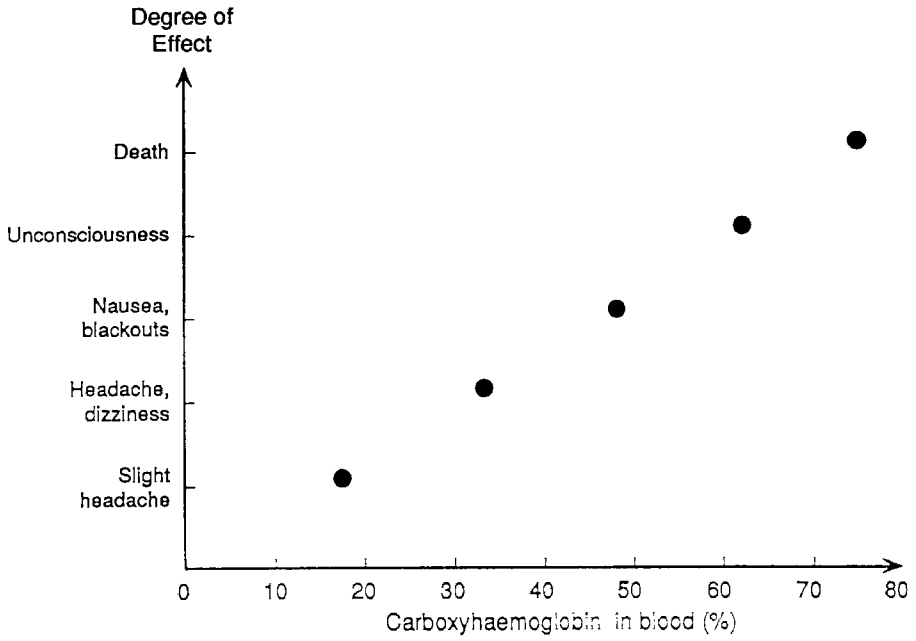
Yet although the scientific community has expressed great interest in applying the technology of molecular biology to epidemiological studies the interpretation of findings has proved difficult. The collection of biological samples from individuals, and the costs involved in such collection, and in assay analysis, have presented problems. The potential difficulties in developing appropriate study designs and identifying willing study populations mean that biomarkers will not be applied widely in epidemiological studies until the feasibility of their use has been evaluated through considerable methodological research.

It should also be noted that the use of biomarkers carry with them a requirement for informed consent. The subject, the epidemiologist and the laboratory technician need to have a clear understanding of the test, its relevance to the investigation, and the ethical context of its use.

Relationship between exposure and effect

Distinguishing between dose–effect and dose–response relationships is crucial to accurate assessment of the effects of an environmental exposure. A dose–effect relationship describes the relation between dose and type or severity of effect in an individual (see Figure 2.3). The severity of the effects following exposure to a toxic agent increases with increasing dose: from physiological adaption to irreversible damage.

Figure 2.3. Dose-effect relationship

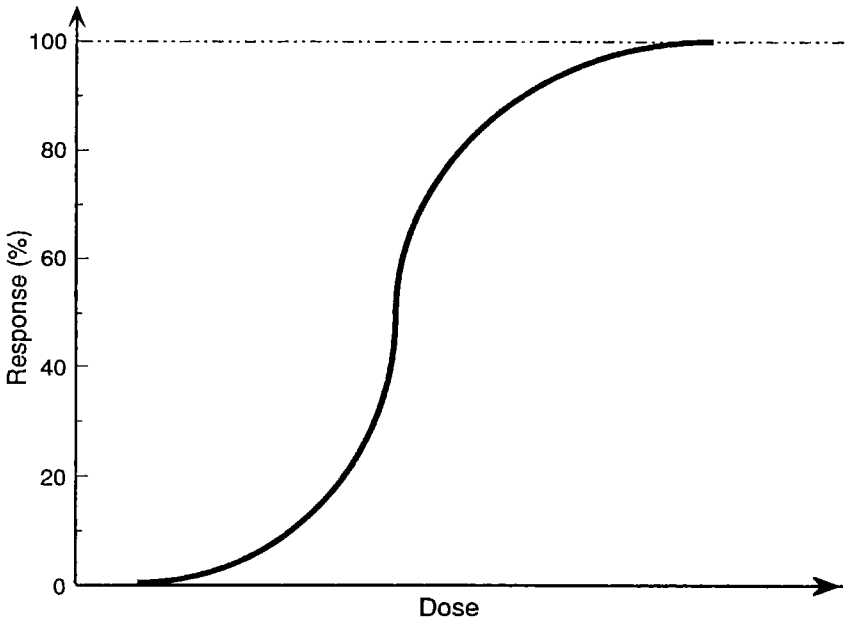


Source: Beaglehole et al., 1993.

Characterizing an effect on the basis of what has been observed in a single individual is generally not sufficient, however, since susceptibility within a population varies. Thus dose–response relationship refers to a population and describes the probability of an effect as a function of the dose. Dose–response relationships are the primary focus of epidemiological investigations.

In general, the dose–response curve shows how the risk of health effect increases with increasing dose. (see Figure 2.4.) In addition, the shape of a particular dose–response curve may indicate the existence of a threshold, i.e. that adverse effects do not occur until a certain dose has been reached. For ionizing radiation, some types of effect, such as acute radiation sickness, develop only if exposure to very high doses has occurred. But a radiation dose below which there is no increased risk of cancer has not been identified, meaning that there is no apparent threshold for the risk of cancer.

Figure 2.4. Dose–response relationship



Source: Beaglehole et al., 1993.

The relation between dose and response may be modified by factors such as age or sex, or by other exposures. For example, children appear to be particularly susceptible to the central nervous system effects of lead (WHO, 1995), and smoking appears to potentiate the induction of lung cancer associated with exposure to asbestos (Hammond et al., 1979) and ionizing radiation (Pershagen et al., 1994a; Pershagen et al., 1994b). Such effect modification can make generalization of the findings of a particular study for application to other populations difficult. Effect modification is discussed more in Chapter 3. Table 3.3 shows an example of the interaction between cigarette smoking and asbestos in causing lung cancer.

As pointed out earlier, the exposures encountered in the general population may be relatively low and the dose-response relationships encountered in a study may be weak. The resolution power of a study can then be increased by selectively including sub-populations with a substantial

proportion of highly-exposed individuals. Focusing on susceptible individuals can also be a means of detecting effects that would be rarer and hence more difficult to observe in the general population. Other options include increasing the precision of the measurements of exposure and confounding factors. As previously indicated, the specificity of the health risk determination is often crucial. An environmental epidemiology study showing a strong and precise dose-response relationship usually gives convincing evidence for preventive action.

2.3 Individuals and study populations

Widespread nature of some exposures

Identifying populations so that a study includes both a sufficient number of highly-exposed individuals, and a suitable non-exposed reference group can be problematic if environmental exposures are widespread. Geographic location or other population-based criteria to define high exposure may not discriminate between people with different exposure levels. Individual exposure measurements including biomarkers may be required. For example, biological markers indicate that some exposure to environmental tobacco smoke is widespread even among people from non-smoking households (Rylander, 1994). The use of a variable such as spousal smoking to estimate exposure may then result in substantial misclassification, leading to a dilution of any apparent association with health effects.

Multicentre collaboration at national or international level is one means of achieving a larger range of exposure and number of subjects with disease than in a study of one geographic area. For example, multinational studies involving several countries in Europe, and covering a wide range of exposures, have been carried out on the acute effects of air pollution on asthmatics. In such instances the study design and all protocols should be agreed upon at the outset so that results are comparable. Pooled analysis of ongoing studies, or meta-analyses of published results may lead to bias caused by differences in study methods. However, problems in interpreting multicentre studies can arise if results appear to differ between locations. Controlling carefully for any extraneous differences between the study sites that could bias the comparisons of the exposure–response relationship is therefore important when undertaking this type of study.

Migration and other sociodemographic characteristics of populations

Migration from rural to urban areas during industrialization has become a major phenomenon in most developing countries (WHO, 1992d). Similarly, significant trans-border migration — due to temporary or permanent relocation of migrant or guest labourers, has occurred in developed and developing countries alike. Since migration can affect health status, problems can arise in environmental epidemiology studies if it is not taken sufficiently into account. Another cause of bias is migration from a polluted area due to health concerns. For instance, if a study was conducted to assess the impact of urban air pollution on asthma induction in children, the association would be underestimated if families with atopy had moved away from the polluted urban areas before the study took place.

In particular, migration can complicate the study of diseases with long latency since the individuals affected may have moved far away from the relevant exposure by the time the associated disease becomes apparent. On the other hand, migration may provide valuable information about the role of environmental factors in disease. Studies have compared disease rates in migrant populations to those in the population from which the migrants came (e.g. Japan) and in the area to which the migrants moved (e.g. USA). Such comparative studies can differentiate between environmental and genetic factors since genetic factors do not change with migration (see Table 2.4).

Table 2.4. Relative risks of mortality from cancer of the stomach, liver and colon among Japanese men in Japan. Japanese immigrants to California, and sons of Japanese immigrants compared with white men in California, aged 45–64 years.

	Relative risk (compared with California whites)		
Cancer sites	Japanese in Japan	Japanese immigrants to California	Sons of Japanese immigrants
Stomach	8.4	3.8	2.8
Liver	4.1	2.7	2.2
Colon	0.2	0.4	0.9

Source: Buell & Dunn (1965).

Lifestyle factors such as tobacco usage, alcohol drinking and dietary habits, are also often strong determinants of health status. It is well known that sociodemographic factors may determine some of the environmental exposures to which an individual is exposed. For example, lead exposure in children in the USA is often higher in urban low-income areas. Type or severity of effect may likewise be influenced by sociodemographic factors. For example, gastrointestinal absorption of lead is higher among children of impoverished families whose diets are deficient in calcium and iron. Thus these children may not only undergo higher lead exposure, but also be more susceptible to the effects of such exposure than children of more economically privileged families.

Sociodemographic characteristics may also act as confounders, as has been found in studies on lead exposure and intellectual development in children. For example, children from families of low socioeconomic status may experience high lead exposure, but the low educational attainment of their parents may be an independent risk factor for their slower intellectual development. In this instance, controlling analytically for the effect of parents' education would be necessary so that the effect of lead exposure on their children's intellectual development could be determined. But controlling the effects of sociodemographic factors when studying environmental exposures can be difficult because these factors may interact with the environmental exposures in complex and indirect ways (see Box 2.3).

Box 2.3. Chagas disease

Chagas disease is a debilitating parasitic disease widespread in Central and South America. Infection with the parasite *Trypanosoma cruzi* occurs via bites of blood-sucking triatomid bugs, which are its natural vector, or via blood transfusions from infected donors. An estimated 18 million people suffer from the disease and some 100 million are at risk. A leading cause of heart disease in several Central and South American countries, it is in many areas one of the chief causes of sudden death in apparently healthy young persons.

Infection most often occurs in low-income families in rural areas where poor housing conditions provide breeding sites for the disease vector. Although considered a rural disease, the rural–urban migration of many rural people infected with Chagas disease has made it a significant cause of morbidity and mortality in urban areas. For instance, although the Federal District in Brazil is considered to be free of triatomid bugs, Chagas disease has been found to be the cause of about one death in ten among people between 25 and 64 years of age. This places the disease roughly on a par with cancer, heart disease, and stroke in terms of contribution to causes of death.

Source: WHO, 1992d.

Sensitive sub-populations

Examples of sensitive sub-populations are people with allergy and other types of hyper-reactivity. Basing studies on such individuals can increase study efficiency. For example, only subjects with asthma might be included in investigations of acute effects of airborne particles.

Specific susceptibility may be attributable to hereditary factors. Heredity appears to be important for some conditions such as allergy, but generally, environmental exposures have a greater influence on disease occurrence. Smokers are often particularly vulnerable to the effects of environmental exposures. This is illustrated by the strong interaction between residential radon exposure and lung cancer that was observed in a recent study. The study estimated that about 80% of radon-related lung cancer occurs in smokers (Pershagen et al., 1994a, 1994b).

Sensitive sub-populations often constitute a considerable proportion of the population. For example, one-third of all children have some type of allergy or hyper-reactivity. Prevention of health effects in such groups is therefore of great public health significance and identification of sensitive sub-populations crucial to the development of effective preventive strategies. Environmental epidemiology can help to identify sensitive sub-populations by examining disease rates among different sub-populations (classified by personal characteristics, such as age) that have comparable environmental exposures.

2.4 Ethics in environmental epidemiology

For an environmental epidemiologist, ethical concerns include the need to protect study subjects and the need to inform them honestly of all possible risks and benefits. Ethical concerns also relate to the design of a study and publication of results.

Epidemiologists are paying increasing attention to the subject of ethics. This is partly due to the recognition that publicly-funded enterprises and agencies — many of which employ epidemiologists — should be held accountable for their activities and interests. That said, it was only recently (the mid-1980s), that epidemiologists introduced the discussion and consideration of ethics to their profession (Soskolne, 1989). Environmental epidemiologists were even later in their acknowledgment of this topic. The International Society for Environmental Epidemiology (ISEE) stimulated discussion of ethical issues when it established a Standing Committee on Ethics and Philosophy in 1991. Considerable attention now focuses on ethical issues that are specific to environmental epidemiology (Beauchamp et al., 1991; Coughlin, 1997; Coughlin & Beauchamp, 1996; Fayerweather et al., 1991; Soskolne, 1993). This is not to suggest that those previously engaged in this area of research or practice were unable to draw on a substantial body of ethics. Indeed, the theory of ethics upon which environmental epidemiology depends is that of utilitarianism, and is common to many other professions. According to this theory, an action should be based upon the principle of securing the greatest amount of good for the greatest number of people.

It should be pointed out that the debate about the development of ethical guidelines for environmental epidemiology has focussed on the protection of the rights of the subjects being studied and the ethical conduct of the epidemiologist in relation to funding and reporting of a study. An issue of equal importance is the ethical responsibility of an epidemiologist to carry out studies of potential health hazards as soon as they have been identified or suspected. A decision not to carry out a study may be unethical, particularly when the data needed for the study is readily available. For example, the data collected in medical surveillance programs in industry could be used to establish new epidemiological knowledge about occupational health hazards, but few of these programs analyze the data in this manner.

Ethical and legal guidelines for research

Many academic and research institutions and government agencies use independent bodies to review the ethical aspects of research proposals that refer to the participation of human subjects (as voluntary participants). In a number of countries, the law protects the rights of individuals to privacy and requires that research participants provide "informed consent". Elsewhere, requirements are less strict or non-existent.

Whatever the circumstances, though, environmental epidemiologists should adhere to the ethical codes for scientific research developed internationally such as those of the Council for International Organizations of Medical Sciences, or adapt them to local circumstances, whenever they carry out research involving human subjects (CIOMS, 1991 and 1993). Ethics guidelines and codes of professional conduct embrace concern not only for individual subjects and public welfare, but also for interpersonal conduct among professionals. For example, the *Ethical Guidelines for Epidemiologists* (Fayerweather et al., 1991) of the Industrial Epidemiology Forum cover:

- obligations to the subjects of research;
- obligations to society;
- obligations to funding agencies and to employers;
- obligations to colleagues.

A statement known as *The Toronto Resolution* has attempted to identify those elements that each profession's code of conduct and/or guidelines should include. The intention is to ensure that professions are cognizant of the social implications of their activities and oppose:

- prejudice;
- activities that cause adverse environmental impacts;
- actions that threaten individual human rights;
- actions that serve militarism.

The Toronto Resolution also stresses that codes of conduct and professional guidelines should:

- explain clearly the principles underlying any specific guidelines;
- refer to measures to ensure the compliance of members;
- establish mechanisms for reporting violations of the code of conduct and/or guidelines for protecting whoever reports a violation;
- anticipate consequences to participants and society from the profession's perspective;
- cover both applied and basic professional activity;
- specify procedures for peer review;
- urge that all basic research results be made universally available;
- urge the dissemination of professional standards in order to encourage "socialization" of students of the discipline.

Confidentiality and privacy

In epidemiology, "confidentiality" refers to protection of the privacy of the individual subjects or participants, and assurance that medical information relating to them will not be disclosed to any third party. Without the medical information, the epidemiological study cannot be carried out, so access needs to be assured in an ethical manner, normally involving consent from the patient. In many clinical investigations, a request for permission to disclose personal information can be made directly to the patient. The situation in other types of investigation may differ. Some epidemiological studies require that only existing or historical medical records of large groups of individuals be examined; no direct contact with individuals is needed for the purposes of the study. In which case, a request for permission to use personal information may be very difficult. Indeed, if individuals have relocated or perhaps died, it may be difficult or impossible to contact them to seek access to their medical records; attempting to do so could increase the cost and complexity of the study considerably. Moreover, if many individuals cannot be contacted, the participation rate may be much lower than for investigations which from their outset are based on contact with individual patients or cases. Such non-participation bias can undermine the validity of a study greatly.

Problems about confidentiality can be difficult to resolve. This is because the needs of a society as a whole may conflict with those of its individual members. An important example is the use of existing medical records to conduct epidemiological investigations. In such instance, the benefits to society from the study's findings must be balanced against the potential loss of privacy of the patients. Even if the type of information that is to be obtained from the patient's medical record is relatively non-controversial (e.g. blood pressure), those handling the data might become privy to more sensitive information (e.g. psychiatric history, tests for venereal diseases,

genetic markers). If reasonable safeguards can be assured, the agreement of the appropriate hospital staff regarding access to medical records may be sufficient. To help ensure that the confidentiality of the medical records is not violated, those responsible for extracting the data can be required to take an oath to the effect that they will not make any disclosures to a third party.

Issues of confidentiality of information may also relate to groups of people or specific populations. For example, it is sometimes necessary to avoid the precise identification of small groups of individuals, such as those who live in a defined area, or who constitute a specific minority. It may also be necessary to protect groups of doctors and nurses who are involved in the care of patients, and sometimes even medical institutions or the region served by them.

General concern about confidentiality can mean that access to medical information relating to individual patients is denied. This may occur even if the identities of patients and doctors have been removed — for example, if the information was originally collected for a different purpose and explicit consent for the currently proposed study was thus never obtained. However, if a physician and an ethics review committee are convinced that obtaining access to medical records is justifiable, this problem may be resolved. Reasonable safeguards to protect the confidentiality of individual medical records can be provided and adequate assurances given that the information will be used only for statistical purposes and, in the interests of the common good. In these instances the personal identifiers are often discarded after exposure and outcome data have been linked for analysis.

Until relatively recently, society was safeguarded not only by the reluctance of many physicians to divulge personal information, but also by the fact that the means of accessing, transferring and linking information were not very sophisticated. Developments in computer technology have changed this picture dramatically and the confidentiality of medical information stored on computerized information systems is giving rise to even more perplexing ethical questions. There is the legal question as to who "owns" the medical records: the physician, the patient or the health authority? And does ownership apply to the paper on which the record is written or to the information itself? Further difficult questions relate to legal right of access and legal right to deny access to this information.

Social responsibility to protect, inform and provide benefits

Environmental epidemiologists have a social responsibility to ensure that the work they undertake is beneficial to the communities and public they serve, and that the benefits of this work far outweigh any risks that may be involved.

Risk–benefit assessments of a proposed study should therefore take socio-cultural factors, such as traditions, as well as considerations of the relevance and importance of the research, into account. If a population does undergo some risk by participating in a study, any benefits that accrue following completion of the research should be available to that population.

Each participant should be informed in detail of the study results that relate to him or her, and of the interpretation of these results. These results must be held strictly in confidence and should not be released, even to a family physician, unless prior written authorization has been obtained from the participant. The participant should have the right to be informed of any adverse medical

conditions concerning him or her that were discovered during the study. An individual patient or subject may benefit directly from participating in a study, if, for example, a previously undiagnosed disease or susceptibility is detected. Alternatively, if no abnormality is found, the subject may be reassured.

Conversely, informing participants that they have contracted a disease that cannot be treated effectively may be counterproductive. Likewise, if no abnormality has been found, the transient and limited value of a negative examination, and hence the possible need for repeat examinations, may not be appreciated. (These factors have been addressed in the area of screening.) Furthermore, job or insurance opportunities could be denied to subjects in whom abnormalities have been detected and reported. Issues relating to providing informed consent are discussed further in Chapter 8.

Policy of openness

The leader of a study team is responsible for making results (with their interpretation) available to the study participants, the public, policy- and decision-makers, and the scientific community. If the study was imperfect or the results would be prone to misinterpretation, the epidemiologist must clearly indicate the unsatisfactory aspects of the study so that the public and decision-makers are not misled. It follows that, as far as possible, all interested parties should be encouraged to contribute to debate concerning interpretation of the study findings.

Once a conclusion has been issued about the relation of a measurement of biological effect to a quantum of exposure, a broader dialogue can proceed. The epidemiologist should play the role of expositor in this dialogue, but recognize that the evidence that is being presented represents only one factor. When discussion extends beyond the realm of epidemiology to include economic or social factors, the epidemiologist can speak only as an ordinary citizen.

Unfortunately, non-scientific factors can lead to a blurring of scientific evidence, and unnecessary concern may result. Examples would be when a study's sponsor attempts to influence the dissemination of results, or when an investigator incompletely discloses study findings in the belief that the stakeholder (or potentially affected) community may be upset about the findings. It would appear that the only way of systematically and correctly avoiding this is by a declared policy of openness on the part of the scientific community and by as much exchange of scientific information as possible. It is to be expected that those responsible for measurements, assessments, and interpretations are committed to ensuring scientific integrity.

Communication with the public

A report of the results of a study must be prepared in precise and accurate scientific language. A simplified report for presentation to policy-makers, the public, and in some cases the mass-media, is also often required. However, since the latter type of report will of necessity omit many technical details, the risk of misinterpretation may be high. Scientists and epidemiologists who were involved in the original study should therefore review and approve any such material. The report presentation may need to make explicit the limitations of the epidemiological approach, and the need for support from additional studies before drawing firm conclusions.

An epidemiological investigation may produce a result based on descriptive or cross-sectional data analysis, but require the corroboration of a follow-up study, or a case–control study. In the absence of the latter studies, the researchers would be faced with the dilemma of whether or not to publish the results of the initial analysis. The public could be alarmed by the initial finding, and so much so that repercussions ensue. For instance, property values may decline in an area in which a particular environmental contaminant has been detected, or people may move away from such an area. On the other hand, withholding the information may be inappropriate. In most societies, it should be possible to explain the strengths and limitations of an initial finding so that any concern is placed in context, pending the outcome of a more definitive study.

2.5 Deciding whether to do an epidemiological study

Frequently in environmental epidemiology, a decision is required as whether to conduct a formal study, or simply to base action on current knowledge. This is often the case when a local disease excess or cluster is observed. If a clear body of evidence is available, an epidemiological study may not be warranted. Often local descriptive data ("body count") is required to convince decision-makers that preventive action is worthwhile.

In many countries, the media and general public have become increasingly interested in the occurrence of time–space clusters of cancer, respiratory disease and congenital malformations. In the USA, there are around one thousand reports of these types of health effect each year. But even for such a developed economy, the in-depth study of all of these reports would be impossible. They would require vast human and financial resources with little anticipated useful outcome (Wartenberg & Greenberg, 1992). The following steps should therefore be taken before proceeding to a full-scale environmental epidemiological study:

1. confirm that human exposure or health effects exist;
2. conduct an initial risk assessment;
3. define the purpose of conducting the study;
4. assess the feasibility of the study;
5. ensure the study satisfies ethical requirements.

Confirming that human exposure or health effects exist

If there is a suspicion that a disease outbreak has occurred, the first priority is to ascertain whether indeed this is happening and to clarify the characteristics of the disease. When the disease outbreak has been confirmed, it remains to be determined whether human population exposure is occurring. These conditions must be satisfied with a reasonable degree of certainty if a full study is to be justified. Two examples illustrate this process. In a Turkish town of 400 people, a physician diagnosed two persons with mesothelioma. Further research covering the previous four years identified a total of 23 deaths from mesothelioma. The potential source of exposure appeared to be a nearby quarry, to which the entire population of the town had probably been exposed. The study of this disease cluster led to the identification of a new carcinogen: the mineral erionite (Baris et al., 1978). A different outcome resulted from investigation of congenital malformations and low birth weight in the San Francisco Bay Area of the USA (Shaw et al., 1992). Population distributions were assessed on the basis of potential sources of indirect

exposure to different ground and water pollutants. Further investigation showed that there was no association and that exposure to specific agents was not significant.

In other situations, the health events or diseases that are occurring may not be unusual, but their incidence may allegedly have increased. Thus epidemiological studies are often requested by communities who believe that potential environmental exposures have led to an elevated rate of cancer incidence, or to health problems such as birth defects. Cases that have led to a perception of increased risk may be identified by community members. Before undertaking a study, the epidemiologist should confirm whether the identified cases are valid and whether the health effects could plausibly be due to the exposures of concern. Existing data on environmental exposures and health effects may be available, but more commonly some effort must be expended to evaluate reported exposures and health effects. Investigators should resist the temptation to embark on a study based purely on community concerns with little or no evidence of probable human exposures or biologically plausible health outcomes.

Conducting initial risk assessment

Clusters generally come to the attention of health officials in one of two ways:

- a concerned citizen or health professional may perceive more than the expected number of cases, frequently in association with a suspected hazard;
- a community requests a review of disease data in relation to a hazard of concern to decide whether an excess risk exists;

To deal with these concerns the environmental epidemiologist would start with an initial risk assessment. Each environmental problem implies a greater or lesser human health risk. Risk assessment is the process by which the form, dimension and characteristics of that risk are estimated (Loehr & Lash, 1990; Aldrich & Griffith, 1993). Therefore, a general assessment of the risk in question should be made before an environmental epidemiology study is embarked upon. Starting with efficient and quick environmental impact methods, a check can then be made, using basic statistical methods, to ascertain if the event of concern has a higher than expected incidence (Neutra et al., 1992; Badii et al., 1994). This is more difficult for rare events (e.g. cancer) than for those that are more common (e.g. diarrhoea or reproductive events).

What questions could a study answer?

Environmental epidemiology studies are undertaken for a variety of reasons, as described in section 2.1.1. The epidemiologist should understand the motivation for conducting a study and evaluate whether the study is likely to address the issues of concern. Studies are most likely to be productive if they are based on clearly stated hypotheses. In most instances, the hypothesis is that an association exists between an environmental exposure and clearly specified health outcomes. Causal associations can be demonstrated most clearly if comparisons can be made between groups that have been exposed to several levels of the environmental agent. If possible, therefore, explicit hypotheses about the form of exposure–response relationship should be evaluated.

The specific objectives to be accomplished should also be identified. The statement of objectives should clarify issues to be addressed, the possible scope of the study, the nature of the study population, and the principal exposures and health outcome variables to be assessed.

Community leaders must be identified and involved in any study from its inception, as part of the research team or steering committee for the research project (ATSDR, 1990). This will ensure that if a study is undertaken it will address the issues of concern to the community. If a dialogue is established between the community leaders and those who are responsible for the study, the elucidation of risk factors and the design of effective intervention programmes will proceed more quickly. Thus community leaders should be consulted even before a decision is made to undertake a study.

Feasibility issues and ethical considerations

Once the potential for exposure and health effects has been established and the study questions have been posed, the next step is to assess whether an epidemiological study would be feasible. This will depend largely on the resources, personnel and existing information available for accessing exposure data, the type of study methods to be used and whether the data are historical or concurrent. In addition, the possibility of identifying early markers of exposure or disease can be helpful. Scarcity of health events or low probability of exposure may preclude the success of the study, or increase its cost and the length of time required to undertake it. The following issues should be considered:

- types of data that need to be collected;
- access to existing data;
- required size of the study population;
- availability of an adequate sampling frame for selecting the study population;
- capability for measuring environmental factors and health outcomes;
- ability to identify and control for confounding factors and effect modifiers;
- availability of personnel and resources.

When deciding whether to undertake a study, the epidemiologist should try to anticipate the range of possible study outcomes and consider how the study findings might be used. In principle, studies that are likely to yield inconclusive findings should be avoided. In many instances, a modest but clearly feasible study will yield more useful findings than a study which attempts to find a definitive conclusion, but which may not be feasible.

The ethical aspects of conducting a study that were discussed above must also be considered. Study subjects must be protected from undue harm, including loss of confidentiality, and informed of possible risks and benefits. The study must do more good than harm from the point of view of the study participants and community as a whole, and should have reasonable potential for addressing issues of community concern. In short, ethical considerations should guide the decision to undertake a study, each step in conducting a study, and in reporting study findings.

Chapter 3

Review of Epidemiological Principles

Learning objectives

This chapter summarizes epidemiological concepts and methods with which the student should already be familiar. After reading this chapter, the student should:

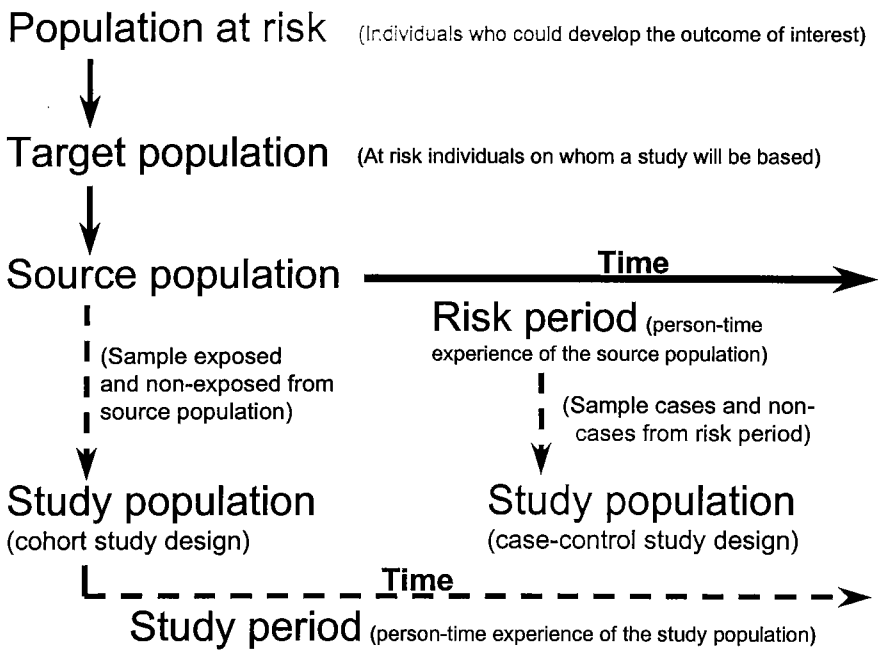
- Be able to define and apply the following measures of health in populations: incidence rate, incidence proportion and prevalence proportion.
- Be able to define and apply measures of association and population disease burden, namely: rate ratio, risk ratio, odds ratio and attributable risk.
- Understand the concepts of precision, random error, and population sampling.
- Understand the principles of effect modification, validity, systematic error, and confounding.
- Be able to identify and explain the criteria for causality in an environmental health context.

3.1 Concepts of populations

A fundamental premise in epidemiology is that disease does not occur strictly at random in the human population. It is possible to identify causes of a disease by studying the patterns of disease occurrence among different populations, in particular, by comparing health outcomes of exposed populations with that of unexposed populations. It is important to understand the concepts of population used in epidemiological studies since human populations are the basis of these studies. Figure 3.1 shows the relationship between the key terms used in this text to refer to populations in epidemiological studies. The **general population** refers to all individuals, theoretically in the world, but generally meaning in some specified area such as a country or region.

The **population at risk** consists of the individuals in the general population who could develop the disease of interest. People who already have a disease are not at risk of developing the disease and, therefore, are excluded from the population at risk. Also individuals who previously had an infectious disease, such as measles, and develop immunity against re-infection are not at risk. The population at risk may also be limited to a single gender for gender-specific diseases such as prostate cancer in males or uterine cancer in females. Even among the population technically at risk of developing a disease, there can be substantial variation in risk so that the most investigators may focus on only a sub-population as the practical population at risk. For example, most studies of breast cancer focus on women even though males can develop breast cancer.

Figure 3.1. Concepts of study populations in epidemiology



It is not necessary and generally not desirable to study the entire population at risk in order to evaluate epidemiological associations between exposure and disease. For example, although the entire population is at risk of developing a heart attack, the risk may be exceedingly small among children. Therefore, an epidemiological study may define a more limited at-risk population. An example would be a study of the occurrence of heart attacks only among persons older than 40 years of age. Populations may also be limited by geography (e.g. persons in a particular region), gender, or other characteristics. The resulting specifically selected population at risk is considered the **target population**. The target population is the population to which the interpretation of the study results will most likely be generalized. For example, the findings of a study regarding the risk of heart attack due to an environmental exposure among a particular study population of 40-50 year old males may be interpreted only in terms of the risk that exists among the target population of all 40-50 year old males.

The **source population** is the population at risk from which the study participants will be sampled. The source population is closely related to the target population, and many investigators consider these to be synonymous terms. However, in practice, the source population is defined in more operational terms than the target population. Continuing with the above example, the target population may be defined as 40-50 year old males living in a particular region of a particular country, but the source population may be limited to those who have access to medical care if they were to develop a heart attack. Understanding the source population is important in an epidemiological study since it is the population to which the findings of a study are applicable. Investigators may subsequently consider biological mechanisms, toxicological data and other information to consider whether the findings of a study may be generalized to the source population, the target population or broader population at risk.

As described below, a fundamental consideration in epidemiological measurements is not only number of individuals who develop a health outcome, but also the rate at which the outcome occurs. Therefore, populations concepts in epidemiology incorporate both the number of persons and the time period of the observation. The **risk period** is the total time during which the individuals in the source population are at risk of developing an outcome of interest. The risk period is quantified as the "person-time" at risk since it incorporates both the number of individuals at risk and the amount of time each individual is at risk. For example, a population of 200 males living at risk over a 10 year period would represent 2000 person-years at risk.

The **study population** consists of individuals sampled from the source population and included in the study. For example, investigators may use a tumour registry to sample males between the ages of 50 and 65 who develop lung cancer within a region over a specific period of time. The source population includes all males between the ages of 50 and 65 living in the region who could develop lung cancer during the specific period of time, the study period consists of the total time during which persons in the source population are at risk during the specific period of time, while the study population includes only the persons actually recruited into the study.

As described in chapter 6, epidemiological study designs differ in the manner in which the study population is sampled from the source population. In a cohort study (see section 6.2.3), a sample of individuals at risk is drawn from the source population. This study population is followed over time to measure their health outcome occurrence. The person-time follow-up experience of the study population in a cohort study is the **study period** (analogous to the risk period in the source population). The study population and study period in a cohort study are directly observable since the individuals in the study are actually under observation. In a case-control study (see section 6.2), the study population consists of a sample of cases and non-cases (or referents) from the risk period, or person-time experience of the source population. The full hypothetical study population in a case-control study is not directly observed, but it is conceptually identical to the study population in a cohort study.

It should be pointed out that not all epidemiologists agree on a standard terminology for all these concepts of population and study base. For instance, you may find the term "base population" in the literature. It refers to the study population in a cohort study (Miettinen, 1985).

3.2 Measurements of health in populations

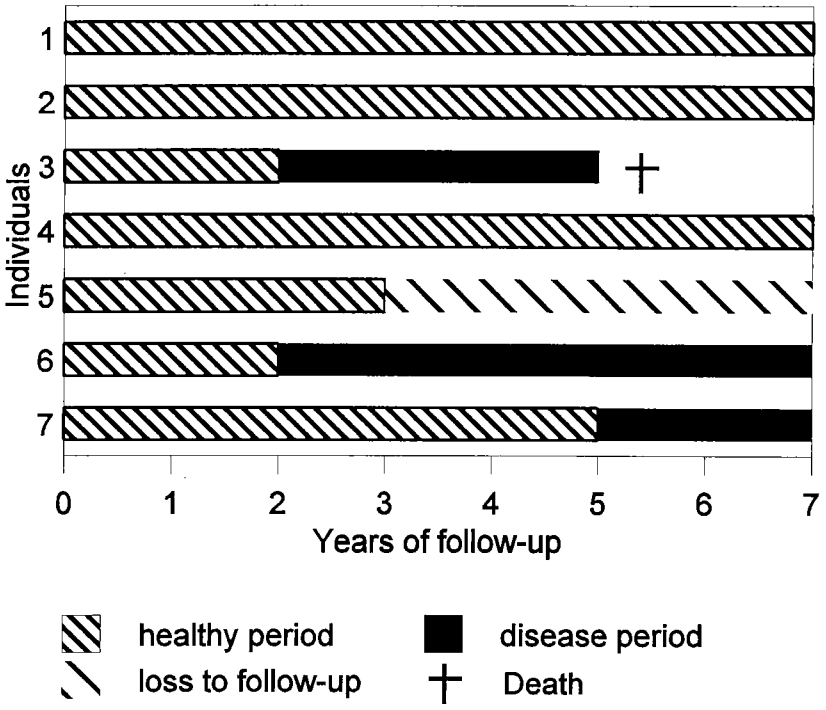
As pointed out in *Basic Epidemiology* (Beaglehole et al., 1993) epidemiological investigations are based on quantification of the occurrence of disease or health outcomes in populations. The most basic measure of disease occurrence is a simple count of the individuals affected. This can be used to show time trends for epidemics and to quantify a specific health problem for a defined geographic area. However, the number of persons who have a disease is generally not informative unless additional information about the population is also available. Exceptions include extremely rare conditions, such as Ebola haemorrhagic fever, for which any case may be so unusual as to warrant attention. In order to investigate disease occurrence, the size of the population at risk and the time period during which the number of cases was observed must usually be known. Disease occurrence can then be calculated as a function of the number of health outcome events per unit size of the population (e.g. cases/1000 people; cases per 100,000 people) for a specified time period or a specified point in time.

Occurrence of disease is expressed as **prevalence**, based on the number of cases that exist at a designated point in time, or **incidence**, based on the number of new cases that occur during a given period. Specific definitions of these disease occurrence measures have been developed according to how the cases and populations are observed.

Incidence measures

Perhaps the most common measure of disease occurrence is the (person-time) **incidence rate** (Miettinen, 1985; Rothman & Greenland, 1998). This is a measure of the disease occurrence per unit time. The incidence rate is based on the notion of observing a population over time to determine the number of individuals who develop the health outcome that is being investigated. In Fig. 3.2 the study period of a hypothetical population is presented. People enter the study in a particular year, and some of them subsequently develop the disease of interest. Some eventually die from the disease, but others from another cause. By the end of the study period, some members of the cohort have died from the disease under study, some are still alive, and others have been "lost to follow-up" because they died from other causes or because contact with them was lost.

Figure 3.2. A hypothetical study population followed over time



Source: Beaglehole et al. 1993.

It is important to emphasize that study participants contribute information only during the period in which they can develop the disease and can be studied. Individuals who cease to contribute information include those who die and who therefore cannot develop the disease and those who emigrate and are lost to follow-up during the study. However, although a person may die and thereafter be unavailable for further follow-up of health status, the timing and cause of death will nevertheless constitute important information. In other words, exposed people may die at a younger age than people who are not exposed. So it is important to consider how much time each person contributes to the study by using the **person-time** method.

Figure 3.2 illustrates how person-time is measured in a small hypothetical population. In the figure, individuals 1, 2, and 4 are observed during the entire follow-up period and, therefore, each contribute 7 person-years at risk of observation. Individuals 3 and 6 were observed for 2 years before developing disease and each contribute 2 person-years. They no longer contribute years at risk once they develop the disease. Individual 5 contributes 3 years of observation before becoming a loss to follow-up perhaps by moving away. The total period at risk of developing the "disease" in this small population is 33 person-years. Because diseased individuals remain at risk of dying (for example, individuals 3, 6 and 7) the study period at risk for mortality in this small population is 43 person-years (7, 7, 5, 7, 3, 7, 7, years respectively).

The **incidence rate** in a study population takes the form:

$$I = \frac{cases}{person\text{-}time} = \frac{M_1}{Y_T}$$

The formulae presented in this chapter are generally consistent with the notation shown in Table 3.1. So as to not present multiple formulae for each measure, the text will present the formula for the total population (e.g. M_1/Y_T). The structure of the formulae for an exposed population and the non-exposed population are equivalent to those shown for the total population. For example, the incidence rate in an exposed population is $I_1 = a/Y_1$; the incidence rate in the non-exposed population is $I_0 = b/Y_0$;

Table 3.1. Notation for data for calculating disease incidence

	Exposed	Non-exposed	Total
Cases	a	b	M_1
Non-cases	c	d	M_0
Study population (persons)	N_1	N_0	T
Study period (person-years)	Y_1	Y_0	Y_T
Incidence rate	I_1	I_0	I
Incidence proportion	R_1	R_0	R
Incidence odds	O_1	O_0	O

Box 3.1. Example of incidence rate

Pearce et al. (1990) undertook a study of New Zealand Navy personnel who had participated in British atmospheric nuclear weapons tests in the Pacific (Pearce et al., 1990). They compared 528 personnel who took part in the tests during 1957–1958 with 1504 men who were also serving in the navy during this time, but on other ships that did not participate in the tests, and who therefore were not exposed. They followed both groups for the period 1957–1987. The 528 test (exposed) participants — of whom 70 died — contributed a total of 13 923 person-years of follow-up. The death rate (the incidence rate of mortality) was therefore 70/13 923 = 50.3 deaths per 10 000 person-years. The 1504 non-exposed personnel — of whom 179 died — contributed a total of 38 937 person-years. The death rate was therefore 179/38 937 = 46.0 deaths per 10 000 person-years.

The **incidence proportion** or **risk** (also known as the cumulative incidence) is another important measure of disease occurrence. This is the proportion of study subjects who develop the health outcome of interest. Since it is a proportion, it is "dimensionless". The relevant time period must be specified, however. An incidence proportion is conceptually the same as the probability (or risk of) an individual developing the health outcome, as estimated by studying a population. When the disease of interest is rare during the follow-up period (e.g. an incidence proportion of less than 10%), then the incidence proportion is approximately equal to the product of the incidence rate multiplied by the period of observation, assuming no loss of follow-up.

The **incidence proportion** in a study population takes the form:

$$R = \frac{cases}{persons} = \frac{M_1}{T}$$

Box 3.2. Incidence proportion

In the New Zealand study described in Box 3.1, 70 deaths occurred among the 528 test participants during the period 1957–1987. The incidence proportion of death was therefore 70/528 = 0.133 (13.3%). Among the 1504 non-exposed personnel, 179 deaths occurred. The incidence proportion was therefore 0.119 (11.9%). (For these calculations, the fact that some people were lost to follow-up during the study period, and therefore ceased to contribute person-years, has been ignored for the sake of simplicity.)

The **incidence odds** is a third measure of disease occurrence (Greenland, 1987a). It is the ratio of the number of persons who experience the outcome (M_1), to the number of subjects who do not experience the outcome (M_0). This measure is rarely used in cohort studies because the incidence rate and incidence proportion are of more interest, but it is used to calculate the odds ratio in case–control studies.

The **incidence odds** in a study population takes the form:

$$O = \frac{cases}{non\text{-}cases} = \frac{M_1}{M_0}$$

These measures of disease occurrence all involve the same numerator: the number of incident cases of disease (M_1). They differ as to whether their denominators represent person-years at risk (Y_T), persons at risk (T), or those who stay do not develop the outcome of interest (M_0).

Prevalence measures

The prevalence of a health outcome in a particular population is based on the number of existing cases at a specific point in time. The **prevalence proportion** of a health outcome in a population is the number of existing cases at a point in time divided by the size of the population. In common with incidence proportion or risk, it is a proportion and therefore "dimensionless", but the relevant point in time must be specified.

The **prevalence proportion** in a study population takes the form:

$$P = \frac{existing\ cases}{persons} = \frac{M_1}{T}$$

Prevalence is a useful measure of the burden of disease in a population because it indicates the size of the population affected. Prevalence is also an appropriate measure for chronic conditions, such as asthma or diabetes, in which the annual incidence of new cases is very small compared with the number of existing or prevalent cases. The prevalence proportion can also be used to estimate the incidence rate in a population if information about disease duration is available. If we assume that the study population is in a "steady state" (which is when the incidence rates and disease duration, exposure and covariate prevalence are stable in the population) and that exposure and disease status are unrelated to the immigration and emigration rates, it can be shown that the prevalence odds are equal to the incidence rate (I) multiplied by the average disease duration, that is: $P/(1-P) = I \times \overline{D}$ (Rothman & Greenland, 1998). The **prevalence odds** is the number of cases divided by the number of non-cases in the population at a specified point in time. Because the prevalence is a proportion representing the number of cases divided by the total population, the prevalence odds is equal to the prevalence proportion divided by one minus the prevalence proportion.

When the proportion of individuals with the disease in a population is small, the prevalence is approximately equal to incidence multiplied by average disease duration: $P = I \times \overline{D}$. This is because the denominator, $(1 - P)$, approximates one if the prevalence is small.

3.3 Comparisons and associations

An epidemiological study generally seeks to estimate the effect of the exposure of interest by comparing disease occurrence in an exposed group with disease occurrence in a non-exposed group. The comparisons result in quantitative estimates of relative risk and risk difference.

Relative risk as a measure of effect

Three principal measures of effect correspond to the three measures of incident health outcome occurrence described above. The measure of primary interest is often the **rate ratio**, which is the ratio of the incidence rate in the exposed group (a/Y_1) to that in the non-exposed group (b/Y_0). If no association exists between exposure and health outcome, and the two rates are therefore identical, the rate ratio is equal to 1.0

The **rate ratio** takes the form:

$$RR = \frac{I_1}{I_0} = \frac{a/Y_1}{b/Y_0}$$

The **risk ratio** (also called the cumulative incidence ratio) is another commonly used effect measure. It is the ratio of the incidence proportion in the exposed group (a/N_1) to that in the non-exposed group (b/N_0). If no association exists between exposure and health outcome, the risk ratio is equal to 1.0 since the two risks are identical.

The **risk ratio** takes the form:

$$RR = \frac{R_1}{R_0} = \frac{a/N_1}{b/N_0}$$

The incidence **odds ratio** is a third type of effect measure. This is the ratio of the incidence odds in the exposed group (a/c) to that in the non-exposed group (b/d). If the outcome is rare over the study period, the incidence odds ratio is approximately equal to the incidence rate ratio. If no association exists between exposure and health outcome, the odds ratio is equal to 1.0.

The incidence **odds ratio** takes the form:

$$OR = \frac{O_1}{O_0} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Box 3.3. Relative risk measures of effect

In the New Zealand study described in Box 3.1, the death rate was 50.3 per 10 000 person-years among the test participants and 46.0 per 10 000 person-years among the controls; the rate ratio was therefore 50.3/46.0 = 1.09. The incidence proportion of death was 0.133 among the test (exposed) participants and 0.119 among the controls (non-exposed); the risk ratio was therefore 0.133/0.119 = 1.11. The odds of death was 0.153 (70/458) for the test participants and 0.135 (179/1325) for the controls. The odds ratio was therefore 0.153/0.135 = 1.13.

The three effect measures (all of which involve ratios) are generally referred to using the generic term **relative risk**. For each, the ratio of a measure of disease occurrence in the exposed group is compared with that in the non-exposed group. The various measures of disease occurrence rely on the same numerators (incident cases), but their denominators differ since these are based on person-years, persons, or survivors (people who do not develop the disease at any time during the follow-up period). If occurrence of the disease during the follow-up period is rare (e.g. an incidence proportion of less than 10%), they are all approximately equal as in the New Zealand study in which the rate ratio was 1.09, the risk ratio was 1.11, and the odds ratio was 1.13.

Prevalence measures of effect

The **prevalence ratio** is the ratio of the prevalence proportion for the exposed to the prevalence proportion for the unexposed. It is an appropriate effect measure for use in circumstances such as an outbreak of illness due to biological food contamination, for which the period of exposure is short. In such circumstances, the prevalent cases represent the entire number of persons affected. This means that the prevalence ratio is equal to the risk ratio over the entire period of the outbreak.

As noted above, if certain assumptions are met, the prevalence odds is equal to incidence times duration. In this situation, the **prevalence odds ratio** is equal to the incidence rate ratio times the ratio of average disease duration in the two groups.

$$POR = \frac{P_1/(1-P_1)}{P_0/(1-P_0)} = \frac{I_1 \times \overline{D_1}}{I_0 \times \overline{D_0}} = \frac{I_1}{I_0} \times \frac{\overline{D_1}}{\overline{D_0}}$$

An increased prevalence odds ratio may thus reflect the influence of factors which increase the duration of disease, as well as of those which increase disease incidence. However, if the average duration of disease is the same in the exposed and non-exposed groups (i.e. when exposure has no effect on disease duration), the prevalence odds ratio estimates the incidence rate ratio. The prevalence odds ratio is often used as the effect measure in a prevalence study, and may be preferable to the prevalence ratio as an estimate of the rate ratio when assessing the relative risk of a chronic health outcome.

Table 3.2 presents the findings of a prevalence study of respiratory symptoms and exposure to automobile exhaust (Nitta et al., 1993). The population examined consisted of 1517 individuals, of whom 570 lived within 20 m of the roadside; of these 56 (9.8%) reported chronic wheezing. The other 578 participants lived 20–150 m from the roadside and 24 (4.2%) of these reported chronic wheezing. Thus the odds of chronic wheezing for those who lived close to the roadside was 56/514 and 24/554 for those who lived further away. The odds ratio was 2.5, which means that those living close to the roadside were two and a half times more likely to respond with respiratory symptoms. The same result can be obtained by comparing the odds of those who lived near the roadside and who suffered chronic wheezing, with the odds of those who did not suffer chronic wheezing. In other words, the ratio of incidence odds comparing exposed on non-exposed is mathematically equal to the ratio of exposure odds comparing cases and non-cases (controls), or ad/bc.

Table 3.2. Analysis of a cross-sectional study of respiratory symptom prevalence and exposure to automobile exhaust using the prevalence odds ratio

Distance from roadside	Chronic wheeze	No chronic wheeze	Odds ratio
< 20 m	56	514	2.5
20–150 m	24	554	

Source: from Nitta et al. (1993).

Rate difference and risk difference as measures of effect

An analogous approach, using the **rate difference** and the **risk difference** rather than ratios, can be used to calculate difference measures of effect. Ratio measures are usually of greater interest than difference measures in epidemiological studies since ratios make assessment of the strength of effect easier. (This is an important criterion when assessing the relationship between an exposure and a health outcome; see section 3.6.) Ratios also have more convenient statistical properties than difference measures and are useful for assessing the possible role of various sources of bias (Pearce, 1989). However, difference measures may be of value in certain circumstances, particularly when evaluating the public health impact of an exposure. For example, difference measures can be used to estimate how much disease may be potentially eliminated as a result of an intervention to reduce environmental exposure. If no association exists between exposure and health outcome, the rate difference is equal to 0.0

The **rate difference** takes the form:

$$RD = I_1 - I_0 = \frac{a}{Y_1} - \frac{b}{Y_0}$$

Equivalently, the risk difference would be equal to $R_1 - R_0 = a/N_1 - b/N_0$.

Box 3.4. Rate and risk difference

In the study of New Zealand participants described in Box 3.1, the death rate was 50.3 per 10 000 person-years among the test (exposed) participants and 46.0 per 10 000 person-years among the controls (non-exposed); the rate difference was therefore 50.3 – 46.0 = 4.3 per 10 000 person-years. The incidence proportion of death was 0.133 in the test participants and 0.119 in the controls; the risk difference was therefore 0.133 – 0.119 = 0.014.

Attributable risk

The effect of an exposure on an outcome can also be expressed by calculating the proportion of the disease incidence among the exposed population that can be attributed to the exposure. Known as **attributable risk**, this measure of effect is based on the assumption that the rate difference (I_1-I_0) represents the incidence of disease among the exposed population that is due to exposure. Consequently, the proportion of the observed incidence among the exposed due to exposure is the rate difference divided by the observed incidence. Furthermore, by dividing each incidence rate in the formula by I_0, the attributable risk can be calculated using only the relative risk measure.

The **attributable risk** takes the form:

$$AR = \frac{I_1 - I_0}{I_1} = \frac{RR - 1}{RR}$$

Estimating what proportion of an outcome in a total population is attributable to an exposure — known as **population attributable risk**—is also possible if the proportion of the total population that has undergone the exposure of interest is known. This measure is estimable because the attributable risk due to exposure among the "not exposed" is zero by definition (since this group is not exposed). The population attributable risk is therefore the attributable risk among the exposed population multiplied by the proportion of the total population that is exposed.

The **population attributable risk** takes the form:

$$PAR = P_e \times AR = P_e \times \frac{(I_1 - I_0)}{I_1} = P_e \times \frac{(RR - 1)}{RR}$$

The population attributable risk can be a useful measure of effect in environmental epidemiology since it provides an indication of the total impact of an exposure on the population. This measure can reveal that exposures associated with a small relative risk of disease can nevertheless have a substantial impact on a population if a large proportion of that population is exposed, as in the case of air or water pollution.

Box 3.5. Attributable risk and population attributable risk

In the study described in Box 3.1, the rate difference was 50.3 - 46.0 = 4.3 per 100 00 person-years and the death rate was 50.3 per 10 000 person-years among the test (exposed) participants. The attributable risk was therefore 4.3/50.3 = 0.086 — in other words, 8.6% of the deaths among the test participants could be attributed to their participation in the weapons testing programme. If records indicated that 5% of the Navy personnel had participated in the nuclear weapons tests, the population attributable risk would be 0.086 x 0.05 = 0.0043 — meaning that 0.43% of deaths among all Navy personnel could be attributed to some of the personnel participating in the weapons testing programme.

Effect modification

Effect modification occurs when the estimate of the effect of the exposure to the study factor is influenced by the level of another factor in the study population (Rothman & Greenland, 1998). An example of effect modification is the difference, due to gender, in risk of breast cancer. A study of environmental risk factors for breast cancer would either be restricted to females or, at a minimum, maintain males as a distinct study group. But the potential for effect modification is not always so obvious. Lately, the possibility that genetic polymorphisms (genetic variation in the population) influence individuals' capacity to metabolize an absorbed toxin, and therefore their susceptibility to environmental exposures, has been receiving attention. This type of genetic variation is relevant to epidemiological studies because it can result in effect modification in which the same amount of exposure to a toxic substance could result in different levels of effect among different genetic sub-groups of the study population. Effect modification due to variation in susceptibility can occur due to environmental as well as genetic factors. For example, gastrointestinal absorption of lead is increased in children with diets deficient in calcium or iron. Therefore, the risk of lead poisoning would be greater among such children than among children with equal lead exposure, but better nutritional status.

Effect modification can also result from biological interaction between multiple environmental exposures, in which the combined effect of the exposures differs from the total effect of the separate exposures. For example, the incidence rate of lung cancer in persons exposed to both asbestos and cigarette smoke is much greater than the sum of the incidence rates of lung cancer in individuals exposed to either asbestos or to cigarette smoke, but not to both. The relative risk of lung cancer following exposure to both asbestos and cigarette smoke is approximately equal to the multiplication of the relative risks attributable to the separate exposures. Table 3.3 shows an example of effect modification (Hammond et al., 1979). The rate difference from asbestos exposure is 470 per 100 000 person-years among smokers and 47 per 100 000 person-years among non-smokers. Evidently, smoking modifies the effect of asbestos exposure when estimated as a rate difference. However, if the rate ratio is used, the relative risk from asbestos exposure is about five times for both smokers and non-smokers. In this case, the risk of the two exposures is approximately multiplicative and there is no effect modification with regard to rate ratios, but there is effect modification with regard to the rate differences.

Table 3.3. Lung cancer mortality rates per 100 000 person-years at risk in a cohort of asbestos workers compared to other blue-collar workers

	Non-smoker		Smoker	
	Asbestos exposure		Asbestos exposure	
	No	Yes	No	Yes
Death rate	11	58	120	590
Rate ratio	1	5.2	10.9	53.6
Rate difference	-	47	109	579

Source: Hammond et al. (1979).

The fact that the joint effect is more than additive has two important implications. In the public health context it indicates that (all other things being equal) preventing asbestos exposure is more important in relation to smokers than to non-smokers; intervention will prevent approximately 43 lung cancer cases in smokers (53.6 - 10.9) for every 4 lung cancer cases prevented in non-smokers (5.2 - 1.0). (Ideally, of course, all asbestos exposure would be eliminated.) In etiological terms, the non-additive nature of the interaction suggests that asbestos and smoking are involved in at least one common causal process (Pearce, 1989). Thus, assessment of effect modification can be important when studying diseases such as cancer that are known to develop through several stages before the clinical onset of the disease.

3.4 Validity (systematic error, bias and confounding)

The goal of any environmental epidemiology study is to obtain as accurate an estimate as possible of the quantitative association between an environmental factor and disease occurrence. The basic strategy for enhancing the quality of an epidemiological study consists of designing the study to minimize the potential for error. Additionally, the data and study should be monitored while the latter is being conducted to assess quality and to facilitate evaluation and adjustment for error when analysing the data. Yet some error is inevitable.

Error is generally classified as random or systematic (Rothman & Greenland, 1998; Beaglehole et al., 1993). **Systematic error** "pulls" the observed effect estimate systematically towards or away from the true value, while **random error** is the divergence, due to chance, of an observation on a sample from the true population value. **Random error** is due to the variability of data with small numbers but can be reduced by doing a larger study. Systematic error is an inherent feature of the study design and the population under study and cannot be reduced simply by doing a larger study (see section 3.5). **Validity** or accuracy is the extent to which the systematic errors are controlled. The term **bias** is used generally to refer to the presence of systematic errors. There are many different types of bias, but three general forms have been distinguished: confounding; selection bias; and information bias (Rothman & Greenland, 1998; Beaglehole et al., 1993).

Confounding

Confounding occurs if the exposed and non-exposed groups are not comparable due to inherent differences in background disease risk (Greenland, 1985; Greenland, 1986). The differences are usually due to individual characteristics such as age, gender or socio-economic background, or to exposure to other risk factors. If no other biases are present, the following three conditions are necessary for a factor to be categorized as a confounder:

- it must be risk factor for the disease in the absence of the exposure under study (it does not have to be an actual cause; it could be a marker for an actual cause);
- it must be associated with exposure in the study population; and,
- it must not be affected by the exposure or disease. (In particular, it cannot be an intermediate factor in the causal pathway between exposure and disease. An intermediate factor is one that is caused by the exposure and which, in turn, causes the disease outcome.) (Rothman & Greenland, 1998).

The bias caused by confounding can be either toward or away from the null value of no apparent association between exposure and health outcome, depending on the relationships between the exposure, confounder, and health outcome. For example, the effectiveness of medical treatment of acute diarrhoeal disease could be evaluated by comparing sick children who received hospital-based medical treatment with children who received traditional home treatments. Better health outcomes might be observed among the children who received the hospital-based treatment. However, these children may live in a city and may be better nourished than the children who received the traditional home treatment. The seemingly greater efficiency of the hospital-based treatment might therefore be spurious. The apparent effect of the hospital-based treatment would be greater than the true effect because of confounding due to nutritional status.

Alternatively, the children who received hospital-based treatment may have done so because they were considered very ill. In which case, the apparent effect of the hospital-based treatment may be less than the true effect because of confounding due to differences in illness severity between the treatment groups. Multiple confounding factors can also occur and may even act in opposite directions, making clarification of the net bias very difficult.

In epidemiological studies, confounding can be controlled in the study design, or in the analysis, or both. Control of confounding at the design stage can be carried out principally by randomization, restriction or matching (Rothman & Greenland, 1998). Randomization, in which exposure or treatment is randomly assigned to subjects by the investigators, is not an option in observational epidemiological studies because the investigator does not assign exposure. Restriction entails narrowing the ranges of values of the potential confounders, for example, by restricting the study to white females in a particular age-group. However, this approach may limit the number of potential study subjects and the amount of information provided by the study. A third strategy is to match study subjects on potential confounders (e.g. matching for age, gender, and ethnicity). Matching in a case–control study is accomplished by measuring the risk factors in the cases and then selecting controls from the base population for whom the same risk applies. Matching can be an effective means of controlling confounding, but it also can be expensive and complicate the analysis and interpretation of study findings. Methods for controlling confounding by study design are discussed further in Chapter 6.

The most common approach is to control confounding during the data analysis. This involves stratifying data into subgroups according to the levels of the confounder(s) and calculating a summary measure of effect that summarizes the information across strata. Alternatively, multivariate analytical methods can be used to model the effect of exposure while adjusting for confounders. In general, control of confounding requires careful use of *a priori* knowledge, together with assessment of the extent to which the effect estimate changes when the factor is controlled in the analysis. Methods for controlling confounding by stratified analysis and multivariate methods are described in Chapter 7.

Selection bias

Whereas confounding is generally due to biases inherent to the study base, **selection bias** arises from the procedures used to select study subjects from the study base into the study population. Thus selection bias is not usually a problem in a cohort study involving complete follow-up, since all of the available information from the study base may be used. However, bias can occur in a cohort study if subjects are lost to follow-up due to factors associated with both exposure and outcome, or if exposure is an effect modifier for the association of study participation with health outcome (Greenland, 1987a). For example, bias could occur in a cohort study of air pollution and lung disease if individuals in polluted areas selectively migrate away from the study area when they develop respiratory symptoms and are consequently lost to follow-up. It is important to recognize that bias in the effect estimate would occur only if loss to follow-up differed between the exposed and the non-exposed; for example, if individuals living in a polluted area with respiratory symptoms were more likely to migrate than non-exposed persons with similar symptoms.

Selection bias is of more concern in case–control studies since these entail sampling from the study base. In particular, selection bias can occur in a case–control study if controls are chosen in a non-representative manner: for example, if exposed individuals were more likely to be selected as controls than non-exposed individuals, as might happen if the controls for lung cancer cases are hospital patients with other diseases (tobacco smoking increases risk of lung cancer and many other diseases).

If appropriate information is available, selection bias can be assessed and controlled using methods identical to those used to assess and control confounding. In particular, selection bias can sometimes be controlled in the analysis by identifying factors relating to subject selection and controlling for them as confounders. For example, persons who have received higher education may be more likely to be selected for or to participate in a study and may have a different disease risk than less-educated persons. If education is negatively or positively related to the exposure of interest, this bias can be partially controlled by collecting information on social class and controlling for social class as a confounder in the analysis. As with confounding, assessment of the likely strength and direction of bias may be possible even if selection bias cannot be controlled for directly. The analytical approaches to controlling confounding and selection bias are discussed in Chapter 7.

Information bias

Information bias results from misclassification of the study subjects with respect to disease or exposure status. Information bias is also commonly called **observation bias, measurement bias,** or **misclassification bias** since information errors commonly arise while measuring or classifying study variables. Thus information bias refers to the individuals included in the study whereas selection bias refers to the selection of the study subjects from the study base.

Two general forms of information bias are recognized, depending on whether the classification error for exposure is independent of the classification error for disease, or vice versa (Rothman & Greenland, 1998). The distinction is relevant because the consequences of the biases are different. **Nondifferential information bias** occurs when the likelihood of misclassification is the same for both groups being compared (Checkoway et al., 1989). In other words, the likelihood of misclassification of disease would be the same among the exposed and non-exposed groups, or misclassification of exposure would be the same among the diseased and non-diseased (well) groups. Of course, in many instances, there is some amount of misclassification in measuring both exposure and disease in a study population so the likelihood of misclassification must be considered for all measured variables. Non-differential misclassification generally biases the effect estimate towards the null value of no apparent effect. So non-differential information bias tends to produce "false negative" findings, and is of particular concern in studies that find no association between exposure and disease. An example of probable non-differential information bias in environmental epidemiology is the use of a person's residence as a surrogate measure of exposure status. For example, the distance of a person's home from a roadway or industrial facility might be used as a measure of potential lead exposure in a case-control study of neurological disease. There would be information bias - equal among cases and controls - if the surrogate distance measured did not accurately reflect true environmental lead exposure.

Differential information bias occurs when the likelihood of misclassification of exposure differs between diseased and non-diseased, or the likelihood of misclassification of disease differs between exposed and non-exposed persons. This type of bias may pull the observed effect estimate toward or away from the null value. For example, in a case–control study, recall of an exposure such as passive smoking in people with a lung disease might differ from that of healthy people. Differential information bias might therefore occur and could bias the odds ratio in either direction. Similarly, people living close to a factory may report respiratory symptoms more often than people living some distance from the factory in question, simply because they are more concerned about a possible link between their symptoms and the factory's emissions.

Several common sources of information bias have been recognized. Some of these, such as recall bias and interviewer bias, are of particular concern because the bias may vary among the study groups so that the net effect of the bias is not predictable (Drews & Greenland, 1990).

Recall or **reporting bias** refers to differential reporting of information by study participants. For example, persons who have developed lung cancer may be more likely to recall exposure to asbestos or radon than control subjects who have not developed cancer. Several studies have found that persons living in communities near hazardous waste disposal sites tend to report higher prevalences of a wide range of symptoms; a pattern that seems consistent with increased symptom reporting among concerned community residents who are concerned about a potential exposure (see, for example, Ozonoff et al., 1987 and Baker et al., 1988).

Interviewer bias refers to bias in the collection or recording of information by study staff. For example, prior knowledge of the subjects' status, may influence how study staff obtain or record information. Thus an interviewer might probe the environmental and occupational history of a lung cancer case more closely than that of a control person who does not have cancer. These biases are more likely to occur in historical cohort, case–control and cross-sectional studies (see Chapter 6) than in cohort studies since an individual's health outcome status may be known before the history of exposure is determined.

Measurement error refers to errors made in measuring study variables. As such, many investigators consider this term to be synonymous with information bias, but it is important to distinguish between the validity of a particular measure of exposure or disease (measurement error) and the validity of a study's effect estimate (information bias). Measurement errors can be systematic, for example, a spirometer that is always incorrectly calibrated, or random, e.g. because of test-to-test variability. Systematic errors reduce a measurement's validity, while random errors reduce measurement precision, but in both instances these errors will result in information bias and reduce the validity of the main effect estimate. Paying meticulous attention to measurement criteria, study methods, study instruments, staff training and study implementation can enhance both validity and precision. In addition, multiple measurements, using the same or different techniques, can be taken to quantify measurement variability. For example, investigations in a study of childhood asthma might obtain several samples of household dust to be analysed for concentration of allergens, such as dust mite. Results from the multiple samples could be used to estimate variability in the measurements.

Similarly, measurements of health outcomes, especially biological processes, must take into account systematic variability in effects. **Intra-individual variability** in effects can occur for some conditions or physiological functions that display distinct cyclical (e.g. diurnal or seasonal) variation. This variability must be taken into account when recording measurements relating to conditions and functions. For example, lung function varies in accordance with a diurnal or daily pattern so that measurements taken for an individual in the morning may differ from those taken for the same individual in the evening. The time at which measurements are recorded can therefore be critical. **Inter-individual variability** in effects is associated with the range of health outcomes observed in a population following its exposure to an agent; "resistant" and "susceptible" persons will be found at the two extremes of distribution.

Strategies to reduce misclassification should first focus on quality control of measurements. Alternatively, multiple measurements should be obtained for all study subjects to obtain an average best estimate of the true measurement. However, repeating exposure measurements may be costly and place an undue burden on the study subjects. In principle, intra- and inter-individual variation can be estimated by repeated measurement for a representative sub-sample of the study population. The information obtained can then be used to adjust the data analysis, to obtain unbiased estimates of effect.

Interpretation of survey findings depends on the accuracy of the measurements. If a reliable "gold standard" is available, direct assessment of the accuracy of measurements may be possible. Inter-laboratory standardization, which commonly forms part of the quality assessment and quality control procedures of larger studies, is another approach to monitoring measurement accuracy.

Box 3.6. Control of measurement error

Measurement error can be minimized by:

- development of explicit, objective criteria for measuring environmental characteristics and health outcomes;
- careful consistent data collection — for example, through use of standardized instruments; objective, closed-ended questionnaires; valid instruments;
- careful consistent use of data instruments — for example, through use of standardized training and instruction manuals, blinding to the extent possible;
- development and application of quality control/quality assurance procedures;
- use of multiple sources of data;
- data cleaning and coding;
- analysis and adjustment, if necessary, to take account of measurement bias.

If measurement error cannot be examined directly, examining the repeatability of measurements — both within observers (i.e. assessing whether the same observer records the same observation on separate occasions) and between observers (i.e. assessing whether two or more observers record the same observations of the same subject) — may be possible. Discrepancies between observations may occur if the parameter of interest is inherently variable (e.g. blood pressure). But if this is not the case, lack of repeatability will imply lack of validity. Unfortunately, the reverse does not necessarily apply: a repeatable measurement is not always valid.

Quality assurance principles

When designing a study, care must be taken to select optimal quality assurance methods for data collection and data management. The study protocol should state how data will be collected and what methods will be used to monitor and evaluate its quality. A written record of all procedures and any changes in implementing study procedures should be maintained throughout the study. Study staff should be trained and then monitored periodically throughout the study to evaluate data collection techniques. Several common approaches exist for improving and monitoring quality assurance in epidemiological studies, including those listed below. Practical strategies for implementing a quality assurance programme are discussed in Chapter 8.

Blind and double-blind assessment

The potential for bias is often reduced if the data collection is carried out "blind", which means that the study staff or subjects are unaware of the exposure or disease classification of the person being studied. This is the case irrespective of whether information is being obtained by interview, physical examination or clinical investigation. For example, in a survey of cataract and sunlight exposure, study staff who are responsible for assessing cataracts should be kept unaware of the level of their subjects' exposure to sunlight. In such studies, withholding details of the study hypothesis from the interviewer/examiner is also advisable. Subjects can be kept "blinded" by describing the purpose of a study in general terms (e.g. environmental factors) rather than stating the specific exposure being studied. Of course, the statement of purpose must be clear enough to allow for "informed consent" to participate. The issues of informed consent are discussed in Chapters 2 and 8.

In a "double-blind" study, both the study staff and the person studied are unaware of the person's exposure classification. For instance, in a study of lead impacts on children's behaviour, a psychologist who assesses the child and interviews the parent will be unaware of the child's blood lead level. At the time of assessment, information pertaining to the child's blood lead level can also be withheld from both the parent and the child. (Information about the child's exposure can of course be given to the parent after the assessment has been made.)

Inter-observer variation

If two or more observers are collecting information, discrepancies may occur with respect to how that information is collected. For example, discrepancies may occur due to the way in which each investigator records information, performs physical examinations or interprets clinical findings such as X-rays or electrocardiograms. Observer performance and interpretation may also vary from time to time and place to place. Inter-observer variation should therefore be monitored regularly and systematically. Sometimes, this type of variation can be measured only by requesting an observer to perform or interpret a standard test at different times, or by evaluating an observer's work at random intervals. For example, in a multi-centre study of environmental factors in childhood asthma, investigations provided centralized training of interviewers to ensure consistency. Subsequently, one interview in ten was observed by a supervisor to verify that the interview was being administered in a consistent manner (Mitchell et al.1997, Kattan et al.,1997).

Intra-observer variation

Measurement variation can even occur in a single observer who obtains measurements at different times under identical conditions. For example, a technician may record slightly different numbers when measuring a subject's blood pressure. Also an interviewer may ask a question with a different tone of voice, which may elicit a different response from the subject's. The main strategy to control intra-observer variation is to train study personnel to be as consistent as possible.

Inter-instrument and intra-instrument variation

All instruments are subject to some variation, as well as to limitations in terms of accuracy. The measurements taken by an instrument may vary over time, or from place to place, due to changes in environmental factors or because of changes in the conditions in which it is operated (e.g. fluctuations in electrical current). Each instrument should be evaluated before use and in any given setting. Changes in barometric pressure, temperature, or humidity, in particular, can affect the functioning of physiological measuring instruments. In addition, the moveable parts of instruments and their individual components can develop defects, or the materials of which they are made may deteriorate. The use of standardized instruments helps to reduce these problems. For example, a study of respiratory function should use an identical model of spirometer for all measurements. Calibrations and frequent checks are required to minimize intra-machine differences: if necessary, adjustments can be made or correction factors determined and applied. The functioning of the instrument, such as its consistency or linearity must be determined. Inter-instrument differences should be noted carefully; such differences can easily introduce bias.

Inter-laboratory and intra-laboratory differences

Quality assurance procedures should be established within a laboratory, and between it and a reference laboratory. They should deal not only with analytical quality assurance, but also, if the study requires use of biological material, with quality assurance during the sampling and storage of that material. Quality assurance checks should be carried out during the study, but also before it starts.

The usage of any laboratory instrument depends partly on whether reference values exist with which results can be compared. Most blood measurements, for instance, have a wide range of "normal" values. The counts cited as abnormal also vary greatly among laboratories. Analytical quality assurance procedures must therefore be followed if routine blood counts are to be used. In this way, the small but nevertheless significant differences between populations that can result from differences in exposure to environmental agents will be detected.

3.5 Precision (random error, statistical power)

As mentioned above, random error is variation in measurements due to chance alone. Random error arises due to biological variability, measurement variability, and sampling variability. Sources of **biological variability** include diurnal variation; changes related to factors such as age, diet, and exercise, and environmental factors such as season or temperature (Greenberg, 1993). Biological variability may also be the result of randomly-distributed susceptibility or genetic differences. **Measurement variability** may be due to inaccuracies in the performance or calibration of a measurement instrument, or misreading or incorrect recording of information from an instrument. Further random error may be introduced following mistakes made during data recording, processing, or management. Biological and measurement variability can lead to systematic or random errors. Even random errors in measurement (lack of precision) lead to systematic errors or bias when using the findings of the measures to derive effect estimates (i.e. due to non-differential misclassification).

Sampling variability occurs because the study participants are always a sample of a larger population (Beaglehole et al., 1993). If a study was repeated on the same source population, some variability would occur each time in the actual study population sampled from the source population. The primary strategy for increasing precision, given variability in the source population, is to increase the size of the study population. Studying large numbers of subjects can be expensive, however. The need for precision must therefore be balanced against considerations of cost and logistics.

An environmental epidemiology study should be large enough to ensure that the estimates of the hypothesized effects will be sufficiently precise, given the anticipated variation among the variables analysed. Statistical analysis of epidemiological studies is discussed in Chapter 7. As noted in that chapter, the aim of statistical analysis is to estimate the precision of observed effect estimates or other estimates. For example, a statistical analysis of a case–control study would typically present the odds ratio as a measure of effect, with a 95% confidence interval as an indication of the precision of the odds ratio estimate. The analysis may also present a p-value as an indication of the likelihood that random error is responsible for the observed association. Although these values cannot be calculated until the data have been collected, estimating the precision of a study by making some reasonable assumptions about the statistical distributions

of the variables to be analysed, is possible. Since the precision of a study is very much influenced by its number of subjects, these calculations can be used to estimate the number of subjects required to achieve the desired statistical power. (See, for example, Lwanga & Lemoshow, 1991 and Breslow & Day, 1987.)

Sampling the population

Studying every member of a population is usually not feasible. Sampling is therefore carried out to enhance efficiency and reduce cost. The quality of the sampling influences not only precision, but also the potential for selection bias and for generalizing the study findings.

The **sampling unit** is the basic unit around which a sampling procedure is planned (Kelsey et al., 1996). In most studies, individuals are selected from a population and then examined for exposure characteristics and health outcome status. The individual is therefore the sampling unit. The sampling unit does not necessarily consist of people; vital events (births, deaths) or individual records in studies based on existing data could also constitute a sampling unit. In ecological studies, the sampling unit is a group or community.

The totality of the sampling units of the population to be studied is called the **sampling frame**, i.e. the population from which the sample is selected. A sampling frame is a concrete listing of or method to access the source population. The sample (study population) is a subset of the sampling frame. Examples of sampling frames include lists of consecutive hospital admissions, community census enumeration lists and randomized lists of telephone numbers for communities in which virtually all households have telephones. It is essential to identify a well-defined sampling frame from which a representative sample can be drawn. "Informal" sampling methods, such as use of volunteers, does not usually result in representative samples.

The prerequisites for a "reliable" sample are:

- **goal–orientation**: the sample selection should be determined by the study objectives (or goals), study design, and local conditions;
- **representativeness**: the sample should be representative of the population to be studied so that inferences drawn from the sample can be generalized to that population with measurable precision and confidence;
- **coverage**: adequate coverage of the sample is essential if it is to be representative; unavailability, a high rate of refusal and loss of follow-up can render a sample unrepresentative;
- **size**: a sample should be large enough to minimize sample variability and to enable estimates of the population characteristics to be made with reasonable precision;
- **cost–effectiveness**: the sample should be such that savings in time and cost can be achieved without unduly undermining the study objectives; certain types of sampling might be more cost-effective than others, depending on local conditions.

Methods for sampling persons from populations have been described in several epidemiology and statistical textbooks (see, for example, Kelsey et al., 1996; Kahn & Sempos, 1989). While a simple random sample may seem the best means of obtaining a representative sample, in practice obtaining stratified, systematic, cluster or multi-stage samples is usually more efficient.

Stratified sampling is more efficient than simple sampling if the distribution of subjects — according to relevant variables, such as age or gender — is not uniform. If the population can be subdivided into sub-groups that are more homogeneous with respect to variables of interest than is the population as a whole, study precision can usually be increased through stratification (Kelsey et al.,1996).

One of the principal advantages of systematic sampling is that determining the entire sampling frame of potential subjects before beginning the sampling is unnecessary. Moreover, under field conditions, systematic sampling is often simpler than random sampling. Thus sampling births in a hospital by making a systematic sample of every fifth birth would be easier than applying a random selection procedure to each separate birth, with a one-fifth possibility of that birth being selected for recruitment. The disadvantage of systematic sampling, however, is that identifying patterns in the sampling frame that could make the systematic sample unrepresentative can be difficult. Selecting sampling intervals that are not related to exposure or health outcome is therefore important. For example, sampling of births in a hospital should not be based on the day of the week since elective induction of complicated births may be more likely on some days than on others. Systematic samples are also limited in that estimating the variance may be difficult unless it can be assumed that a systematic sample from a randomly ordered population would not differ from a simple random sample. Most studies make this assumption and analyse systematic samples as if they were simple random samples.

Many environmental characteristics or health outcomes are relatively rare. Furthermore, identifying unbiased sampling frames of individuals for environmental epidemiology studies can be difficult. Studies that use more complex, multi-stage sampling approaches to enhance efficiency are therefore likely to become increasingly common (Morgenstern, 1993). However, maintaining a clear understanding of the relationship between the actual study subjects and the base population using these complicated sampling schemes, is a considerable challenge.

Box 3.7. Definitions of sampling methods

Simple	Each sampling unit in the population has an equal chance of being included in the sample.
Stratified	The population is divided into strata, or groups of sampling units that have certain characteristics in common, and a random sample of units is drawn from each stratum.
Systematic	The selected sampling units are spaced regularly throughout the sampling frame beginning with a randomly selected unit (e.g., every fifth birth in a hospital).
Cluster	Clusters rather than individual sampling units are first selected from the population and observations are then made on all individual sampling units within the selected clusters.
Multi-stage	Primary sampling units are selected from a population. Secondary sampling units are then sampled from each primary unit, and so on. Multi-stage sampling is similar to cluster sampling, but additional sampling is carried out within the clusters.

3.6 Criteria for causality

An epidemiological study generally aims to determine whether an environmental exposure is causally responsible for disease occurrence. A causal association is one in which a change in the frequency or quality of an exposure or characteristic results in a corresponding change in the frequency of the disease or outcome of interest. However, demonstration of a valid statistical association between exposure occurrence and disease occurrence in a population is not sufficient to conclude that the association is causal; additional criteria should be considered. Hill (1965) elaborated a systematic approach to causal inference, and an updated version of this approach is described in Table 3.4.

The **temporal relationship** is crucial; the cause must precede the effect. This is usually self-evident, but difficulties may arise in studies (usually case–control or cross-sectional studies) when measurements of exposure and effect are made at the same time (e.g. by questionnaire, blood test). In order to evaluate whether the temporal relationship is logical, the induction period (the time period between the causal exposure and disease initiation) must be understood for the disease being studied. For some conditions, such as lung cancer, the induction period may be 15 years or more. Therefore, for carcinogenic agents, the causal exposure must precede clinical disease by at least this induction period.

An association is **plausible** if it is consistent with other knowledge. For instance, laboratory experiments may have shown that a particular environmental exposure can cause cancer in laboratory animals, and this would make the hypothesis that this exposure could cause cancer in humans more plausible. However, biological plausibility is a relative concept. Many epidemiological associations were considered implausible when they were first discovered but subsequently confirmed by experimental studies. An example is the occurrence of severe osteomalacia in a cadmium polluted area of Japan (WHO, 1992a). Initially, the mechanism of cadmium-induced bone toxicity was not known and the cause was subject to much dispute. Lack of plausibility may simply reflect lack of medical knowledge.

Consistency is shown if several studies give the same result. If a variety of designs are used in different settings, the likelihood that all studies are making the same mistake is minimized. However, a lack of consistency does not exclude a causal association. In certain studies, the exposure levels may be so low that no health effects occur.

The **strength of association** is important in that a strongly elevated relative risk may be more likely to be causal than a weak association that could be influenced by confounding or other biases. However, the fact that an association is weak does not preclude it from being causal; rather, it means that excluding alternative explanations for the observed association is more difficult.

A **dose–response relationship** can be said to exist when changes in the level of exposure are associated with changes in the prevalence or incidence of the effect. The demonstration of a clear dose–response relationship provides strong evidence for a causal relationship since it is usually unlikely that confounding or other biases would produce a consistent dose–response relationship.

Table 3.4. Guidelines for causal inference

Temporal relation	Does the cause precede the effect? (essential)
Plausibility	Is the association consistent with other knowledge (e.g. regarding mechanism of action, evidence from experimental animals)?
Consistency	Have other studies had similar results?
Strength	What is the strength of the association between the cause and the effect? (relative risk)
Dose–response relationship	Is increased exposure to the possible cause associated with increased effect?
Reversibility	Does the removal of a possible cause lead to reduction of disease risk?
Study design	Is the evidence based on a strong study design?
Judging the evidence	How many lines of evidence lead to the conclusion?

Source: Beaglehole et al., 1993.

Reversibility is also relevant in that when the removal of some possible cause results in an observed reduction in disease risk, the likelihood that the association is causal is strengthened. However, the health effect of exposure may be irreversible (e.g. total deafness caused by gunshot noise), and the reversibility is not a necessary criterion for causation.

Finally, the causal inference is strengthened by data from studies using good **study design**, and when **evidence** from several different types of studies is available.

Chapter 4

Exposure Assessment

Learning objectives:

- Be able to describe the concepts of exposure and dose.
- Be able to describe the major methods for measuring environmental exposure in epidemiological studies.
- Be able to describe the major methods for modelling environmental exposure in epidemiological studies.
- Understand how to apply the different methods of total exposure assessment, personal sampling and biological assessment in environmental epidemiology studies.
- Appreciate the difference between inter-individual variability and intra-individual variability when interpreting exposure measurements.
- Appreciate the importance of standardization, quality assessment and quality control, with respect to obtaining and analysing environmental measurements.
- Be able to describe the major pathways and media for environmental exposures.

4.1 Concepts of exposure assessment

Exposure to an environmental agent is defined as any contact between a potentially harmful agent present in an environmental medium (such as air, water or food) and a surface of the human body (such as the skin, or the lining of the digestive tract or respiratory tract) (Sexton & Ryan, 1988; Armstrong et al., 1992). Potentially harmful agents or hazards in the environment include chemical substances, physical phenomena such as noise and radiation, and living organisms such as bacteria and fungi. Safety hazards and psychosocial hazards are also important environmental hazards, but this chapter focuses on chemical and physical factors as teaching examples.

Classification, measurement and modelling are the principal methods for assessing exposure. Classification refers to the identification of subgroups of subjects according to ordinal categories of exposure, the most common classification being dichotomous — i.e. exposed and not exposed. Exposure measurement generally relies on the use of an instrument to measure the value of an exposure variable. Exposure modelling uses mathematical models to predict the value of an exposure variable. The models are based on knowledge of the factors that determine or influence the exposure variable, and of the quantitative relationship between these factors and exposure.

Accurate and precise measures of exposure are crucial if the effects of environmental hazards on human health are to be estimated with minimal bias and maximum efficiency. However, the cost of exposure measurement generally increases in parallel with its increasing accuracy and precision. Conversely, if the study size is reduced because the cost per measurement is high, the

precision of effect estimates may be decreased. The epidemiologist must therefore decide how to use available resources optimally, particularly since funding for environmental epidemiology studies is generally limited.

Since obtaining individual exposure measurement is difficult and costly, **surrogate exposure variables** — which are approximations of the actual exposure — are often used. These variables may be derived from questionnaire data, existing exposure data or other indicators, such as distance from a pollutant source. Use of the surrogate variables may make a study more feasible in economic terms, but typically yields lower quality data than when direct individual measurements are used. The epidemiologist must therefore appreciate the impact of choices regarding exposure assessment on cost, feasibility, validity and precision.

Exposure pathways

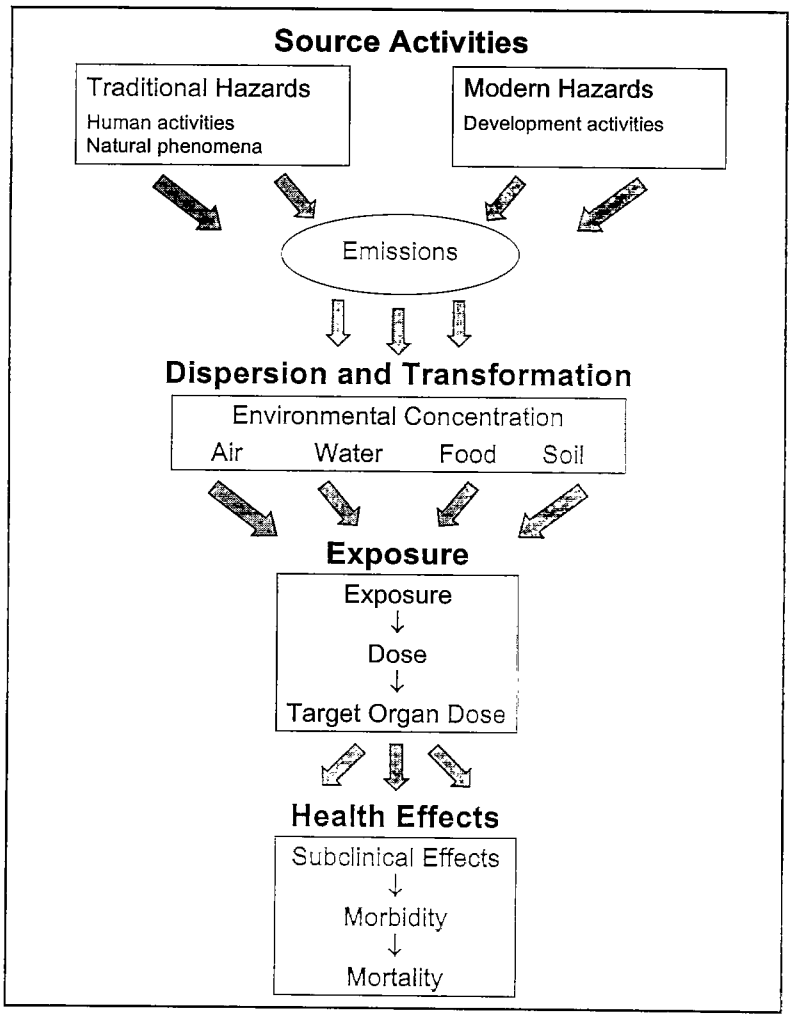
Although environmental health hazards take many forms and can result in human exposure through multiple pathways, the environmental health hazard pathway is broadly similar for each (see Figure 4.1). The starting point in the case of "pollutants" is some form of human activity, or, more rarely, a natural process which releases potentially toxic, infectious or otherwise hazardous agents (pollutants) into the environment. The process of release is termed **emission**. Once emitted into the environment, pollutants typically undergo a process of **dispersion**, during which they are transmitted through the environment via air, water, food or soil — generally referred to as **environmental media**. Exposure occurs when humans encounter the pollutants in the environment.

During dispersion pollutants may undergo various types of transformation. Dilution occurs due to admixture with the transporting medium (e.g. the air or water). Sorting and segregation of pollutants can occur depending on their physical properties - i.e. size, mass or density. Chemical reactions occur, breaking down the original pollutants or converting them into new compounds, which may be more harmful than the original compound. As a result, many pollutants show complex patterns. This complexity means that measuring or modelling exposure patterns, and thereafter determining levels of human exposure can be difficult (Briggs et al., 1996). The epidemiologist should be prepared to measure environmental exposures consisting of multiple potentially toxic agents dispersed via multiple environmental media, with exposure to humans potentially occurring via multiple routes (Armstrong et al., 1992).

Environmental concentration, exposure and dose

The basic distinction between environmental "concentration", "exposure" and "dose" has already been described (see section 1.3, page 9, & section 2.2, page 24). A **concentration** of an agent in an environmental medium refers to the agent's presence in that particular medium, expressed quantitatively. But human **exposure** to that agent occurs only if physical contact takes place between the medium and the body surface of human beings. This distinction is particularly important in relation to environmental media such as soil, which may be polluted, but which do not necessarily come into contact with the human body.

Figure 4.1. The environmental health hazard pathway



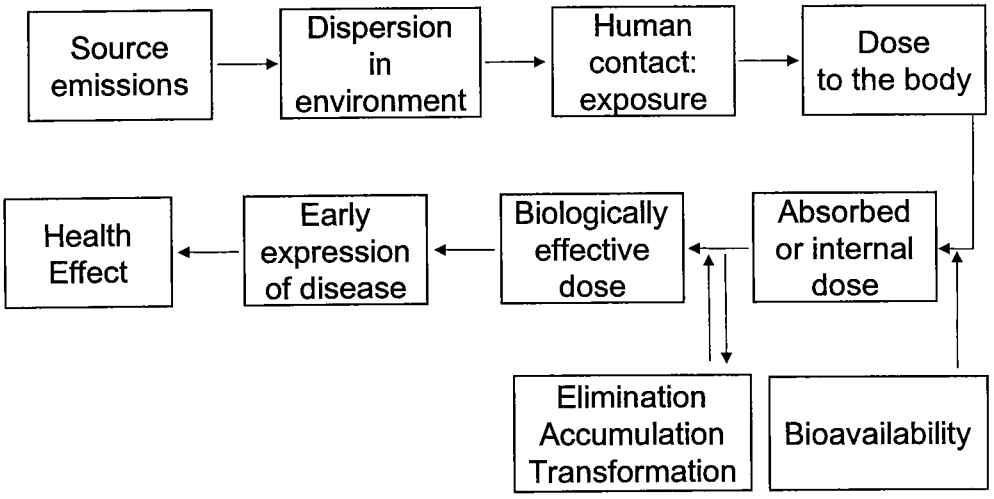
Source: Briggs et al., 1996.

Human contact with the environment, and thus exposure, occur principally through **inhalation** in the respiratory tract, **ingestion** in the gastrointestinal tract, and **absorption** through the skin. The degree of contact between the environmental medium and these body surfaces influences the level of exposure. A major factor influencing the degree of contact is the duration of time in contact with the environmental medium. Human activity may also affect the degree of contact. A person exercising heavily or doing heavy physical work may inhale more than ten times more air per minute than a person at rest. Similarly, the heavily exercising person would consume much more drinking–water (or other liquid) than a person at rest, particularly in hot environments. However, the type of liquid consumed will influence exposure. For instance, some people drink mainly soft drinks, and may therefore hardly be exposed at all to drinking–water contaminants. Likewise, people who spend a lot of time sunbathing or outdoors experience greater exposure to ultraviolet (UV) light than people who never sunbathe and spend little time outdoors.

The term **dose** refers to the quantity of environmental hazard that actually enters the human body. For inhaled or ingested agents, a distinction may be made between the dose that enters the human body and the **absorbed** or **internal dose** that is actually absorbed across the lining of the lungs or gastrointestinal tract. This distinction occurs because some of the agent may be expelled from the body without being absorbed. Measuring dose directly is often difficult. Estimates may have to be relied upon instead. If environmental exposures are complex or difficult to quantify, use of biological markers of the dose (generally called "biological markers of exposure") may be preferable, depending on cost and feasibility. Measuring lead in blood, for instance, is a more effective means of assessing integrated exposure to lead that is occurring via several media than measuring lead in the individual media and then trying to cumulate the individual measurements. (See Chapter 2, page 25, for a discussion of biomarkers.)

Once absorbed into the body, an agent may be distributed in various tissue or metabolically transformed before it reaches the target organ (i.e. the tissue where the agent exerts its harmful effect.) As defined in Chapter 2, **target organ dose** refers to the amount of the agent that reaches the organ where the effects may occur. A similar term is the **biologically effective dose** which is the dose that is directly related to the biological effects. The biologically effective dose is the most relevant measure in assessing health effects because it represents the amount of agent that is biologically active in the target organ(s). However, it is essentially not possible to measure this dose so it must be estimated. Techniques such as dosimetric modeling and pharmaco-kinetic modeling, combined with use of exposure measurements and biomarkers, can be used to estimate the biologically effective dose (Hatch & Thomas, 1993). Figure 4.2 is a more detailed version of the hazard pathway, demonstrating the continuum from human exposure to health effect.

Figure 4.2. Hazard pathway from environmental dispersion to human health effect



Source: Lioy, 1990.

The differences between environmental concentration, exposure and dose can be illustrated using carbon monoxide (CO) as an example. CO is an odourless, colourless gas produced by incomplete combustion of fossil fuel. Automobile traffic is a major source of CO. The concentration of CO in ambient (outdoor) air is an indicator of potential exposure. However,

many people spend considerable time indoors, where concentrations of CO may be lower. Indoor concentrations of CO, on the other hand, may be elevated in the presence of a smoke-leaking stove or header, or smoke from tobacco. So actual exposure may not be similar to measurements of ambient air. The term "CO exposure" should therefore be used to refer to the air CO concentration with which a subject actually comes into contact.

The dose is the amount of CO that enters the body during the time period of interest. After CO has been inhaled, it is transported across the alveolar epithelium (gas exchange lining of the lung). CO in alveolar air reaches equilibrium with CO in blood after about 20 seconds. Thus the concentration of CO in exhaled breath (measured after the subject has inhaled for about 20 seconds) is generally a measure of absorbed dose. CO exerts its harmful effects because it binds more readily to oxygen than does haemoglobin. The carboxyhaemoglobin (COHb) concentration in blood is therefore a measure of the biologically effective dose. Table 4.1 illustrates the relationship between the external CO concentration and the COHb level, according to the physical activity level of the exposed subjects. The table shows that longer duration of exposure at a lower concentration will result in approximately the same COHb as a short duration of exposure at a higher concentration. It also shows that the biologically effective dose (COHb level) will be higher for those engaged in heavy work than for those at lower activity levels, because of the higher respiratory rates of the former.

Table 4.1. Predicted carboxyhaemoglobin levels for subjects engaged in different types of work

Carbon monoxide concentration	Exposure time	Predicted COHb level for those engaged in:		
(mg/m^3)	(minutes)	sedentary work	light work	heavy work
115	15	1.2	2.0	2.8
57	30	1.1	1.9	2.6
29	60	1.1	1.7	2.2
11.5	480	1.5	1.7	1.7

Source: from WHO, 1987.

Similar considerations apply to other exposure pathways such as food and water. Vegetables that are grown alongside a major road may contain lead, but people will be exposed to this lead only if they consume those vegetables. Thus dose refers to the amount of lead actually taken up by the human body from the alimentary canal — itself determined by factors such as nutritional status — following initial consumption. Equally, the presence of nitrates in groundwater does not constitute an environmental exposure unless the groundwater is used as drinking-water. The dose would be determined by the extent to which the nitrates consumed in the water are taken up from the gastrointestinal tract.

Considerations of time and space

Exposure estimation must take into account the **time relationship** between the exposure and effect. For each agent a characteristic time pattern exists relating to exposure and the onset of the adverse health effect(s) (generally known as the latency period). Host factors such as age, gender and immune status may affect this pattern. Some health effects are extremely acute and occur just seconds after the exposure. Examples include annoyance following exposure to odour or noise, or irritation of mucous membrane following exposure to tobacco smoke. Other health effects occur a few minutes after exposure, as with allergic reactions to animal dander and pollen, and asphyxiation following high exposure to CO. Yet other effects occur only after years of accumulated exposure, as with kidney tubular damage due to cadmium exposure. Most cancers develop many years after exposure to the relevant carcinogens. The measurement of exposure therefore requires an understanding of the mechanism of the effect that is being investigated.

The **temporal pattern** of exposure is also significant because it may influence the impact of exposure on the health outcomes. Animal experiments have shown, for example, that the effects of a dose of nitrogen dioxide (NO_2), administered over a fixed period, depend largely upon the mode of administration (for example, continuous intermediate concentration or alternating periods of low and high concentrations) (Miller et al., 1987). Similarly, human experiments have shown that ozone (O_3) exerts a greater effect if administered in a non-constant concentration rather than a constant concentration, over a given period (Hazucha et al., 1992). The same applies to the annoyance effects following exposure to odorous substances. A common practice in measuring exposure patterns is to distinguish between average exposure and peak exposure (which represents the highest exposure concentration even over a short period of time).

Pollution levels typically show a number of different patterns on different temporal scales. In many cases long-term trends exist, reflecting underlying changes in the rates of emission (as a result of technological changes, for example). Annual variations may be superimposed upon these, reflecting year-to-year differences in climate or source activity. Many pollutants also show marked seasonal, weekly and diurnal patterns, due to cycles of activity and short-term climatic and other effects. Short-term pollution episodes may also occur as a result of sudden, accidental releases. Exposure measurements must therefore be appropriate to the time when sampling is carried out and the duration of individual measurements (Briggs et al, 1996).

Considerations of **spatial variation** are also important: human beings cannot come into contact with environmental hazards if spatially separated from them. However, the geography of environmental contamination is complex. Different pollutants may be derived from a wide range of sources. Emissions from these sources may occur either via controlled pathways such as a stack or discharge pipe, or as "fugitive" emissions, which are leaked inadvertently into the environment. Once in the environment, many different processes can act to disperse or transform the pollutants. As a result, spatial exposure patterns can differ markedly: some pollutants may be widely and relatively uniformly distributed, others show more localized patterns. Atmospheric pollutants such as sulphur dioxide which are emitted primarily from the tall stacks of power stations and other combustion plants, may become widely dispersed. Organic pollution of drinking water, in contrast, commonly occurs at the level of a neighbourhood or household. Food contamination can be specific to a particular product and affect all population groups consuming the product, or it can be specific to a household where food storage hygiene is inadequate. Exposure to electromagnetic fields can vary strikingly over short distances.

It is important to be aware of typical spatial exposure patterns, particularly in regard to the household or neighbourhood since many environmental investigations are based in these settings. For chemical and biological agents in particular, the indoor air concentration in the home is often more important — in terms of human exposure — than the concentration in outdoor air. This is because many people spend much more time indoors than outdoors. Similarly, exposure to air pollutants may occur only if homes or workplaces are built on or near sites contaminated with volatile substances and indoor air becomes contaminated.

Integrated and total exposure

Exposure estimation can be relatively straightforward if the study objective is to determine the effect of exposure to a single agent occurring via a single pathway. For example, a study could aim to evaluate whether exposure to salt in the diet is associated with high blood pressure. Various techniques (described below) could be used to measure the salt content in each component of the subject's diet and then to sum the salt content for the components. This approach may be justified if exposure through other pathways is either very low, or if some knowledge of additional pathways already exists.

However, even assessment to a single agent via a single pathway must account for spatial and temporal variation. For example, exposure to carbon monoxide can differ quite markedly between indoors and outdoors. Carbon monoxide exposure will also be affected by the presence or absence to cigarette smoke and use of fuels for cooking or heating. Exposure assessment of CO should therefore measure CO concentration and duration for each of these settings and then estimate the integrated potential exposure — generally using a time-weighted average approach, in which the CO concentration in each setting is multiplied by the proportion of time spent in each setting (see Figure 4.3). Each setting – such as the home or automobile – may be considered a different microenvironment. The concept of microenvironment is discussed further in Section 4.3, pp. 92. This integrated approach also needs to take into account activity level (such as sitting or exercising) which would affect respiratory rate and thus actual exposure to CO in air.

Box 4.1. Example of total exposure assessment for single hazards

Single pathway — Sulfur dioxide (SO_2)

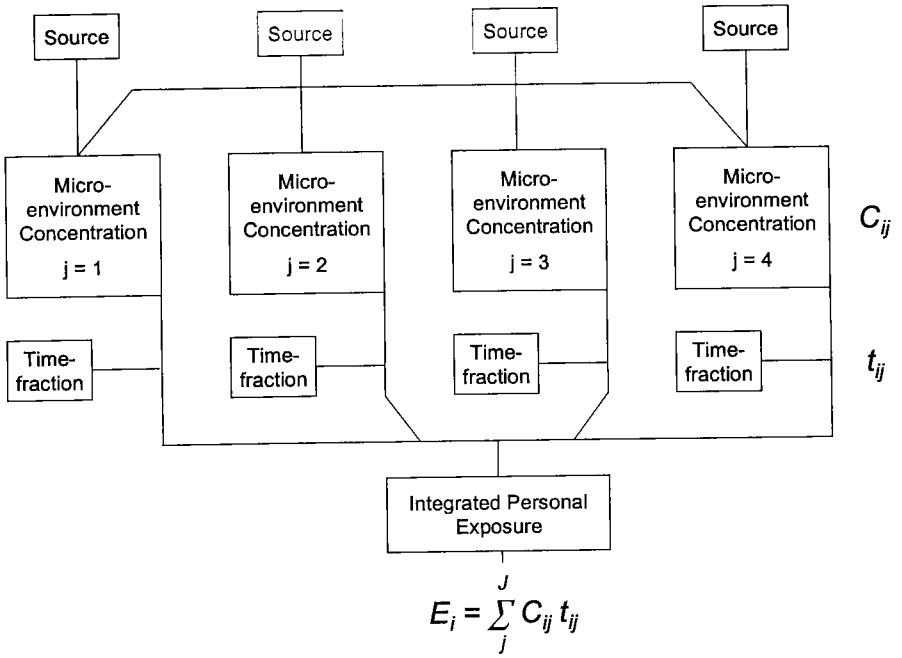
air	100%

Multiple pathways — Benzene

	Intake (µg/day)	Average percent for a pack/day smoker
air	30–300	18.6%
water	1–5	0.3%
food	100–250	17.5%
smoking	600/pack	63.6%

Source: Wallace, 1987.

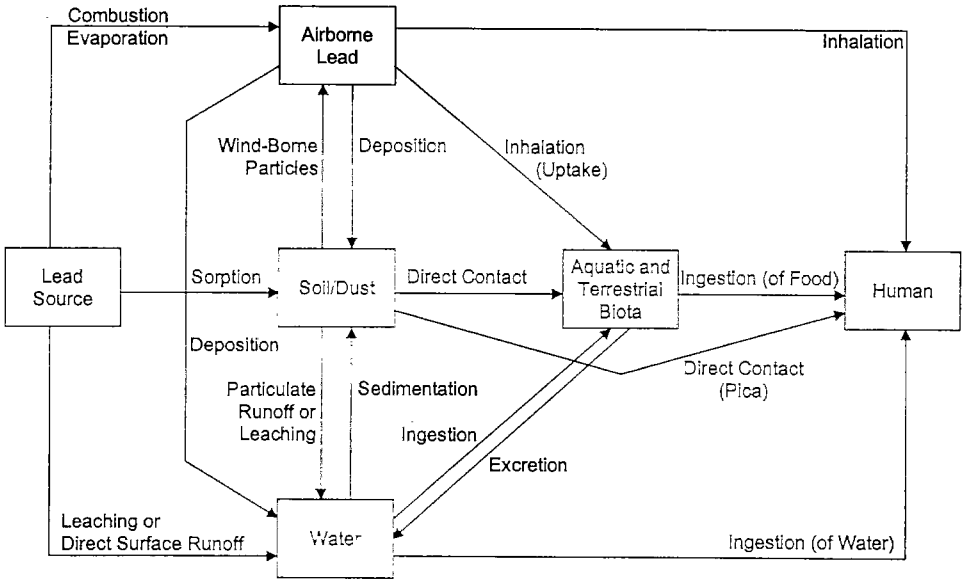
Figure 4.3 Calculation of integrated exposure



More generally it is necessary to estimate exposure that occurs via multiple pathways and in different spatial and temporal patterns. In such cases, **total exposure** must be estimated. For example, exposure to most of the contaminants found in drinking-water and soil can also occur via other pathways. Once an appropriate period for measuring exposure (based on knowledge of the latency and temporal pattern) has been identified, data on exposure concentrations and duration of exposure for each concentration must be collected for each pathway. These data are then combined or **integrated** to obtain the measure of total exposure. The time period during which integration is carried out is called the **time integration period**. Alternatively, a biological marker, such as lead concentration in blood or bone, which in itself represents the integrated exposure for the time period of interest, may be available. Figure 4.4 illustrates the interrelationships between sources and pathways of airborne lead; if exposure via one pathway changes, exposure via other pathways is also likely to change. For example, if the lead content of gasoline is reduced, not only will atmospheric lead concentrations decline, but also the lead concentrations of various other media will decline. The integration process will vary according to the objectives of the study, the agent, and the exposure media, as shown below for air, water and food.

The objectives of a particular study will determine what type of total exposure assessment should be carried out. For example, if the aim is to investigate the relationship between lead exposure and certain health effects, assessing total exposure by biological monitoring, i.e. by measuring the concentration of lead in blood, or in other biological media such as deciduous teeth or bone, would be more efficient than assessing total exposure by environmental monitoring. But if the aim is to assess the contribution of different environmental sources or pathways to exposure levels, exposure should be assessed for each of those pathways.

Figure 4.4 Interrlationships between sources and pathways of airborne lead



Source: adapted from Grisham, 1986; Van Hook, 1978.

Combined exposure

A health effect may be caused by more than one environmental hazard. If so, each of the relevant hazards should be included in the exposure assessment. The objective could be to determine the combined effect of all exposures on the health outcome. In this case, an index of total exposure may have to be calculated. Alternatively, determining the relative effects of the separate or combined exposures may be important. This latter objective addresses the issues of confounding and effect modification, which are reviewed in Chapter 3. Questions to be asked will include:

- What are the separate (independent) effects of each agent or factor?
- To what extent does exposure to one agent or factor modify the effects of exposure to another agent or factor? (See example in Box 4.2.)
- To what extent does combined exposure to more than one agent or factor lead to additive, sub-additive or super-additive effects? (A **sub-additive** effect means that the combined effect is less than would be expected from simple combination or addition of the separate effects. A **super-additive** effect means that the effect is more severe than would be expected from the simple addition of the separate effects.) (See example in Box 4.3.)

Examples of super-additive effects include the increased effect of exposure to an airborne irritant during the allergy season (i.e. when tree or grass pollen is in the air causing allergic reactions) among persons who suffer seasonal allergies. Conversely, exposure to an airborne irritant may modify the effect of an allergen exposure. In a study by Molfino et al. (1991), asthmatic patients who were allergic to ragweed (*Ambrosia* spp.) were exposed to ozone in a chamber. They were more susceptible to a ragweed antigen challenge after ozone exposure, than after exposure to clean air.

Box 4.2. Meteorological conditions and air pollution

Meteorological conditions and air pollution can affect the same health endpoints. Periods of extremely high or extremely low temperature increase mortality. However, periods of very high or very low temperature often coincide with increased air pollution concentrations. Investigations of the effects of air pollution on mortality therefore commonly adjust for temperature effects (Schwartz & Dockery, 1992a; Schwartz & Dockery, 1992b; Dockery et al., 1993a; Schwartz, 1993). Without such adjustment it might be tempting in some situations to conclude that air pollution exposure alone increases mortality, when in fact extremity of temperature is also responsible for some of an observed increase in mortality.

The situation becomes even more complicated if climatic or meteorological conditions and air pollution interact — the effects of air pollution on human health may be increased or decreased when temperatures are extremely high or low. For instance, interactions may occur in relation to air pollution and mortality (Katsouyanni et al., 1993). Or frail human subjects who are stressed already by very high or very low temperatures may experience negative health effects at lower air pollution exposures than they would if temperature stress was absent. Thus the effects of air pollution on the lung function of asthmatics have been shown experimentally to increase after exposure to cold air (see, for example, Sheppard et al., 1984; Dosman et al., 1991). Additionally, the effect of sulphur dioxide on the lung function of asthmatics has been shown to be less under conditions of high temperature and humidity, than when the subjects were studied at room temperature or in conditions of low relative humidity (Linn et al., 1985).

Box 4.3. Measuring exposure to volatile organic compounds

We are exposed to many volatile organic components (VOC) in ambient and indoor air, usually at low concentrations. Indoor sources can include, for example, paints, solvents, cleaning agents, and office products. The total exposure assessment methodology (TEAM) study was developed to assess personal exposure to a mixture of contaminants including VOC (Wallace, 1987). Personal samples were obtained by requesting study subjects to wear small, portable samplers that collected VOC on adsorbents. The study also measured VOC in exhaled breath to obtain estimates of internal dose. The results of personal sampling and of the analysis of VOC in exhaled breath correlated well. However, VOC levels in exhaled breath did not correlate with VOC levels in ambient (general outdoor) air. This showed the importance of distinguishing between indoor exposure and outdoor exposure for this class of substances.

In indoor air, total VOC concentration is the sum of the concentrations of the individual pollutants that make up VOC. Even though the adverse health effects of the individual pollutants have been documented (see, for example, Molhave, 1991), a question remains about their interaction. Specifically, it is unclear whether the VOC pollutants combined cause subtle health effects that would not be anticipated from an assessment of the individual pollutants. Molhave (1991) reviewed the effects of mixtures of VOC on various health endpoints in groups of subjects found in controlled exposure studies to be sensitive to VOC. The studies indicated that sensory irritation and discomfort can be documented at total concentration levels of 3–25/mgVOC/m^3, which is below the concentration predicted from studying the individual chemicals, suggesting that super-additive effects do indeed occur.

Group and individual variability in the relationship between exposure and biologically effective dose

For an epidemiological study, variability among individuals in their absorption, metabolism, and internal distribution of a hazard can lead to random error in measuring the relationship between exposure, biologically effective dose and health effects. Additionally, systematic error due to confounding or effect modification can occur if study groups differ from each other with respect to their toxicokinetic characteristics. This would be the case for example, if the study groups being compared displayed genetic differences in their ability to metabolize an absorbed toxin. Identifying and measuring the relevant genetic characteristics and then using these variables to control for differences among the study groups during the analysis may be possible too.

Social and other factors may also act to distort or mask the association between exposure and health outcome. Climate and the degree of urbanization influence the amount of time people spend indoors or outdoors, and cultural differences in diet, food and water consumption patterns can affect the level of exposure to hazards occurring via water, food and soil. Such factors should be taken into account when developing exposure measurement techniques and when modeling exposure and dose on the basis of environmental concentration measurements.

4.2 Exposure pathways and media

Most pollutants are transmitted through the environment via air, water, soil, dust or food before coming into contact with the human body. The pollutants usually have to be inhaled or ingested in order to result in exposure. Skin contact is also of importance in relation to certain hazards.

Air

Air pollution differs from other forms of environmental exposure in that, once pollutants are present in air, exposure cannot be easily avoided. For instance, since people must breathe, a large population will be exposed if high levels of air pollution are occurring in a city. Nevertheless, levels of air pollution may vary markedly even at the local scale, especially in the case of low-level emissions (e.g. from road transport), due to variations in wind speed, wind direction, rain and geographic topography. Short-term variations in pollution levels will also occur due to variations in emission activity.

Concentrations of pollutants in ambient air may vary considerably over time even if emissions are constant — due to meteorological factors such as wind speed and direction, turbulence, and height of the mixing layer. If the quantity, height and location of emission sources are known, dispersion modeling can estimate air pollution concentrations at specific locations. These models tend to be unreliable, however, if the integration time is short, or if sources are not well defined, or if topography is extremely varied. Such conditions are common in urban areas.

In addition, levels of exposure to outdoor air pollutants will vary depending on the proportion of time people spend outdoors, and the capacity of the individual pollutants to enter the indoor environment. Many people spend much of their time indoors, principally in their own homes. Time is also spent at work, at school and in transit. In colder climate countries especially, the

average amount of time spent outdoors is typically only a few hours or less. The average amount of time spent outdoors also tends to be lower if a country's urban population is much larger than its rural population.

The concentration of indoor air pollution may be lower or higher than the concentration of ambient outdoor air pollution. It may be lower if pollutants are filtered when they pass through the building "envelope". Suspended particulate matter, for instance, is partly filtered out when passing from outdoors to indoors. The extent of filtration depends on particle size: in general, large particles penetrate less than small particles. The concentration of indoor air pollution may also differ from that of outdoor pollution if the pollutants in question react with indoor surfaces. Ozone, for instance, is a chemically highly reactive substance and indoor ozone concentrations are typically much lower than those outdoors. CO, on the other hand, is relatively inert, and in the absence of indoor sources of CO, indoor CO concentrations do not differ greatly from those outdoors.

Indoor air pollution levels may also be much higher than those outside if there are indoor sources of pollution. Many sources of air pollution are found in the home. Unvented or inadequately vented combustion appliances for space and/or water heating, or for cooking, are a major source of indoor air pollution. In many developing countries, the burning of coal or biomass fuels indoors results in concentrations of particulates, sulfur oxides, nitrogen oxides and/or carbon monoxide, that are much higher than concentrations encountered outdoors. In many rural areas, where ambient air pollution is low, the use of biomass fuel in unventilated houses leads to pollution concentrations that are much higher than those in even the worst polluted cities. This form of exposure tends to affect women and children more severely than men since the former spend the longest time indoors (Chen et al., 1990). Many other sources of air pollution have been identified, such as building materials (radon) and soil, particle board and certain forms of cavity wall insulation (formaldehyde), tobacco smoking, pets, fungi and dust mites (allergens).

Ventilation methods and systems should also be taken into account when measuring indoor air pollution exposure because the ventilation rate can be a determining factor of indoor air pollution levels. If no indoor sources of air pollution are present, a low building ventilation rate (low exchange of air between outdoors and indoors) will generally signify lower concentrations of indoor pollution. This is because ambient pollution "decays" in passing from outside to inside. If, however, indoor sources of air pollution exist, a low ventilation rate will increase the level of concentration since less indoor pollution will be transferred from indoors to outdoors.

Exposure assessment may also be complicated if a dwelling has more than one room and pollutant concentrations vary considerably between rooms. The absence or presence of unvented pollution sources, differences in ventilation, and differences in location with respect to local outdoor pollution sources such as roads, are some of the factors that might contribute to such variation. In which case, information about pollutant concentrations and/or pollutant sources in the different rooms, and about the length of time spent in those rooms, must be obtained if a complete exposure assessment is to be made.

Finally, exposure to air contaminants may vary due to physiological factors such as body size, metabolic rate and physical activity which each influence the amount of air inhaled. Thus the amount of air inhaled per person per day ranges from less than 10 m^3 to more than 30 m^3. Minute volumes of inhaled air may range from less than 10 l/m during rest to over 80 l/m during heavy

exercise. Table 4.2 is taken from *Reference Man* (ICRP, 1975) and illustrates the correlation between breathing volume, age and workload. This type of basic physiological data is of great value when exposure estimates must be calculated for different age, gender and ethnic groups.

Table 4.2. Selected lung ventilation rates at different levels of activity as a function of age

		Minute Ventilation Rate (litres/minute)		
Subject	Weight (kg)	Resting	Light activity	Heavy work
Adult male	68.5	7.4	29	43
Adult female	54	4.5	16	25
Adolescent male (14–16yrs)	59.4	5.2		
Adolescent female	56	4.5		
Child male (10–11yrs)	36.5	4.8	14	
Infant (1yr)		1.4		
Newborn	2.5	0.5		

Source: adapted from ICRP, 1975.

Table 4.3 summarizes the parameters to be considered when measuring exposure to airborne contaminants.

Water

Water pollution is a serious problem in many areas of the world, irrespective of the level of economic development (Yassi et al., 1998). The main pathway of exposure is contaminated drinking-water. Most drinking-water is obtained from groundwater or surface water and can become contaminated with biological, chemical or physical agents. Biological pollution is often of greatest concern, particularly in less developed countries and in rural areas. According to the United Nations Programme of Action from the Rio Conference in 1992, an estimated 80% of all diseases and over one third of deaths in developing countries are caused by consumption of contaminated water (UN, 1993). Diarrhoeal disease due to pollution of water with faecal contaminants is widespread and a major cause of infant deaths. Chemical pollutants in water include nitrates and nitrites, pesticides, volatile organic compounds, heavy metals such as arsenic and lead, and to a lesser extent, mercury, cadmium and other metals (WHO,1997).

For drinking-water pollutants, the time integration period may be short, as with infectious agents, or long, if substances such as carcinogens are involved. John Snow's study of cholera incidence in London was an archetypal epidemiological study of exposure to water contaminants. He showed that the risk of cholera in London was related, among other factors, to the consumption of water supplied by a particular company. Exposure assessment was relatively simple since many people obtained water from just a few pumps.

Table 4.3. Airborne contaminants — parameters for calculating exposure

1. Concentrations (e.g. $\mu g/m^3$)
   a. Microenvironments
   b. Personal

2. Patterns of exposure
   a. Intensity "episode" concentrations versus normal (average) levels
   b. Frequency and duration of contact

3. Transport
   a. Dispersion and advection
   b. Other meterology-related removal rates (washout, fallout)
   c. Indoor ventilation and removal rates

4. Chemistry
   a. Formation rates
   b. Transformation rates

5. Deposition rates ($\mu g/cm^2$)
   a. Environment
   b. Lung

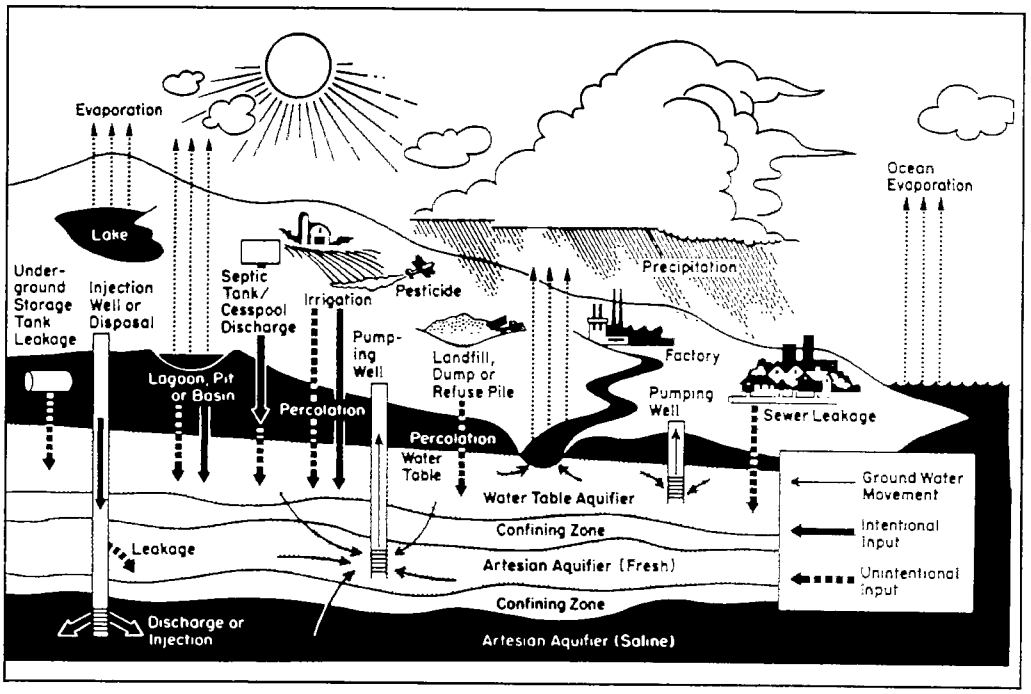
6. Contact
   a. Inhalation (dependent on exercise regimen)
   b. Dermal deposition and permeability
   c. Ingestion (e.g. following food or soil deposition)

Source: Adapted from Lioy, 1990.

The time period of exposure can be affected by retention of the pollutants in the environmental media. In the case of groundwater, turnover rates may be extremely slow — often in the order of hundreds or thousands of years — so that the water remains contaminated for a long period of time. In the case of surface waters, pollutants are often stored in sediment, and released slowly. As a result, contamination often continues long after the original source of pollution has been removed.

In most developed countries, at least in residential areas, drinking-water is piped directly into homes. In many less-developed countries though, most of the population obtains its drinking-water from wells or streams. Such circumstances must be noted when assessing exposure to contaminants in drinking-water. Drinking-water is often supplied from groundwater sources, for instance, and contamination of groundwater will then be of especial concern. Figure 4.5 illustrates various ways in which groundwater can become polluted (Coniglio et al., 1992).

Figure 4.5. Ways in which groundwater can become polluted



Source: Coniglio et al., 1992.

Contamination of drinking-water can occur for a variety of reasons. The raw material may not have been purified adequately, or the purification process (chlorination, for example) may have added pollutants such as chloroform. Piping material can also be a source of contaminants. In many countries lead was formerly used to make drinking-water pipes. If such pipes are still in use and acidic drinking-water passes through them, they may leach large quantities of lead into the water supply. The lead concentration in the water will depend on the length of time that the water has been in contact with the surface of the pipes. The lead concentration of the first water to be drawn in a day is therefore typically much higher than the lead concentration of water drawn later. Various pollutants may also enter the water supply if pipes leak or are permeable to certain pollutants.

Pollutants such as chlorination by-products that are added to or formed after the water has left the pumping station are more difficult to monitor than pollutants that were present in the original water supply. So when assessing exposure, the pollutant concentration should be estimated or measured for water at the time when it leaves the tap. However, if surface-, ground- or well-water is consumed directly, exposure assessment should concentrate on the source itself, and on the contaminant concentration.

Exposure to contaminants in water can also occur if beverages and foods are prepared using contaminated drinking-water or through dermal absorption during activities such as swimming, showering, bathing and cooking. Contaminants can also become volatilized during these activities and then inhaled.

Differences in exposure to water contaminants can also result from variation in personal habits and physiology. For instance, consumption of drinking-water generally varies between one and two litres a day per person and may consist of tap water; bottled water, or a combination of the two. Sufficient detailed information must be collected, therefore, when assessing exposure to pollutants in drinking-water. Interviewing subjects is one means of accomplishing this.

Table 4.4 summarizes the parameters to be considered when measuring exposure to water contaminants.

Table 4.4. Water contaminants — parameters for calculating exposure

1. Concentration (µg/L)
   a. Tap water
   b. Water uses
   c. Effluent — industrial, commercial, residential, uncontrolled dumps

2. Patterns of exposure
   a. Drinking
   b. Cooking
   c. Bathing and showering
   d. Swimming
   e. Laundry

3. Solubility and volatility of contaminant

4. Transport
   a. Groundwater
   b. Surface water
   c. Domestic supply

5. Chemistry
   a. Formation rates
   b. Transformation rates
   c. Degradation

6. Biology (for biological hazards)
   a. Infectious dose or virulence
   b. Organism viability in water

7. Contact rates (µg/L/time) via exposure route
   a. Ingestion
   b. Skin
   c. Inhalation (volatilized)

8. Absorption
   a. Gastrointestinal tract
   b. Dermal deposition and permeability

Source: adapted from Lioy, 1990.

Soil

Soil may be chemically contaminated with a wide range of pollutants, including pesticides and heavy metals such as lead or cadmium. Agricultural activities, industry, landfill and vehicular emissions are all important sources of soil pollution. Soil may also natural concentrations of some harmful chemicals, as well as biological hazards such as helminths.

Direct exposure to soil pollution occurs if soil particles are ingested or inhaled. Ingestion can occur if soil particles adhering to food crops are not fully removed before consumption. In young children, the "mouthing" behaviour that they often exhibit between one and four years of age, can also lead to ingestion of soil pollutants. In normal children, the quantity of soil or dust particles ingested in this way is between 50 and 100 mg per day (Calabrese et al., 1989; Wijnen et al., 1990). (The tendency to eat non-edible objects is known as "pica".) Evidently, children may be at risk if they live and play in a polluted environment. Children may also be exposed to infectious organisms via dust.

Indirect exposure to polluted soil can occur if contaminants are released into the air, taken up by food crops and livestock, or leached into drinking-water supplies or into surface waters. Many pollutants may remain bound to soil particles for considerable periods of time. Gradually, however, they are likely to be leached from the soil into groundwaters or surface-waters. Others are taken up by plants and may enter the food chain. Predicting the environmental fate of soil pollutants is a complex process and usually requires that measurements are taken for each of the relevant pathways.

Table 4.5 summarizes the parameters to be considered when measuring exposure to soil contaminants.

Food

Contamination of foodstuffs may occur as a result of the deposition of pollutants from the atmosphere, through the use of contaminated irrigation water, through application of pesticides and other substances used to cultivate crops or livestock, and through contamination occurring during processing and distribution. Major concerns in these cases include contamination with biological agents or chemicals. Microbiological contamination of food is a major cause of diarrhoeal disease, in part due to contaminated water used to prepare the food as mentioned above (WHO, 1993a; UN, 1993). Additionally, as described in Chapter 1, metals such as mercury and cadmium have caused major poisoning outbreaks; for example, Minamata disease and Itai-Itai disease (WHO, 1990, 1992a, 1992c).

Assessments of exposure to contaminated food must consider all potential sources of exposure. For example, individuals may consume contaminated food and water — even though local food and water are clean — if either are brought in from outside the area. "Ready food" may consist of a mixture of food derived from various sources. If food has been cooked, toxic contaminants may have been eliminated, become more concentrated or even added to the food.

Table 4.5. Soil and dust contaminants — parameters for calculating exposure

1. Concentrations (µg/g)
   a. Dusts – outdoor, indoor
   b. Soil – airborne deposition, resuspension back into the air, landfills, uncontrolled dumps

2. Biology (for biological hazards)
   a. Infectious dose or virulence
   b. Viability of organisms in soil and dust

3. Patterns of exposure
   a. Frequency and duration
   b. Intensity of contact

4. Percolation rate
   a. Soil composition
   b. Water table
   c. Solubility
   d. Transport

5. Volatilization
   a. Contaminant
   b. Soil composition
   c. Top soil and cover

6. Contact rate via exposure route
   a. Gastrointestinal tract — normal population, abnormal ingestion behaviour (pica)
   b. Dermal deposition and permeability
   c. Lung

Source: Adapted from Lioy, 1990.

Assessment of exposure to contaminants in food and water must also consider individual habits in food preparation, and in the choice of various foods and drinks. In practice, however, reliable and representative data are difficult to obtain, particularly if the population consumes a heterogeneous diet, if family units are not very uniform, or if the same element is distributed throughout many items of the diet. Within a population, cultural habits and availability are the factors that most influence the choice of food and drinks. Total exposure is a function of concentration, amount, frequency of intake, and duration (Willet, 1990). In relation to food, the term "daily intake" is often used instead of the term exposure.

Various approaches can be used to assess exposure that occurs via food. To make an overall assessment of dietary intake of toxic elements, information must be collected about types and quantities of the food and drinks that are consumed, and that are representative of national consumption patterns, or those of subgroups within the population. Guidelines for the study of dietary intakes have been developed by the WHO (1985).

Three groups of consumer may require special attention when assessing exposures that occur via food:

- individuals (e.g. infants or the elderly) whose food consumption patterns differ from those of average adults;
- individuals whose metabolism differs from that of ordinary adults (e.g. infants, since they normally absorb lead from the gastrointestinal tract at a higher rate than adults);
- individuals exposed to an above-average concentration of toxic chemicals in the diet (e.g. tuna fishermen whose diet consists largely of tuna fish and who are consequently exposed to high levels of methyl mercury).

Table 4.6 summarizes the parameters to be considered when measuring exposure to contaminants in food.

Table 4.6. Food contaminants — parameters for calculating exposure

1. Source of contamination
   a. Naturally occurring contaminants
   b. Airborne deposition
   c. Water supply
   d. Fertilization and pest control
   e. Waste dumps
   f. Preparation and cooking techniques

2. Biology (for biological hazards)
   a. Infectious dose or virulence
   b. Viability of organisms
   c. Natural toxin

3. Concentrations (µg/g)
   a. Plants, vegetables and fruit
   b. Milk
   c. Drinks and water-based foods
   d. Cooked foods

4. Patterns of exposure
   a. Rate
   b. Frequency
   c. Origin of food — home grown, commercial, local farms, processed

5. Contact rate
   a. Gastrointestinal
   b. Inhalation (during cooking only)

6. Absorption through gastrointestinal tract

Source: adapted from Lioy, 1990.

Physical factors — noise and radiation

Noise

In highly industrialized societies, noise occurs almost everywhere. The major sources of noise are road and air traffic, construction, industry, and people. It is particularly widespread in the production, transportation, construction and mining industries. It can also be high during some agricultural activities. Non-occupational exposure to noise occurs more extensively in urban than in rural environments, although large motorways and airports may contribute considerably to the noise exposure of rural communities. Exposure to noise also occurs during recreational activities (e.g. shooting, motorcycling) and in public places such as discotheques.

In epidemiological studies, total noise exposure should be assessed for each day. Sound-level meters, impulse noise meters and personal dosimeters are used to quantify exposure to noise. All measurements should be conducted and calibrations performed according to accepted standards, such as those of the International Organization for Standardization (ISO). Frequency and duration are essential parameters. Some equipment measures frequency and duration directly, as with frequency weighting networks built into sound-level meters; for other equipment, calculations based on a knowledge of the time pattern are required. To obtain a complete history of noise, tape recorders would be taken to a site (workplace or community). The tapes would be later analysed in a laboratory.

Environmental noise is complex (Rylander, 1992). Various methods of calculating an average have been developed, such as the noise pollution level, the average day and night level and the **equivalent sound level** (Leq) for different parts of the day (Yassi et al., 1998). Each of these techniques also involves calculating a weighting of various sound frequencies. Even so, the average noise level may be difficult to interpret because a few events with high noise level will have the same average as a large number of events with a low noise level. Thus, noise can also be characterized by the peak (or highest) noise level, or it can be described statistically as the time during which a certain level was exceeded (for example, L10 would mean that the level was exceeded for 10% of time). In studies of the effects of aircraft noise, contour noise level lines can be computed for areas around an airport. Exposure of the general population can then be expressed in terms of location of homes (and communities) on a contour line map.

There is no strong documented evidence that environmental noise generally can cause long-term hearing damage (Yassi et al., 1998). However, noise is still an important concern because it may affect health by inducing nuisance. Nuisance may be specific, such as when a conversation is interrupted by a passing aircraft, but it can also be non-specific and give a general feeling of annoyance, discontent, or even fear. Therefore, it may be important to measure the quality of the noise and perception of the noise, for example by using questionnaires or personal diaries, in addition to making objective measurements of noise frequency and intensity.

Ionizing radiation

Ionizing radiation is either electromagnetic (e.g. X-rays or γ-rays) or corpuscular (e.g.α-particles or ß-particles) and emitted by radioactive elements (radionuclides) (ICRU, 1980). Exposure can occur in several settings: domestic, occupational, local or regional. A number of radionuclides are of natural origin and therefore always present in the environment. These, together with contributions from cosmic rays, constitute background radiation. Total background radiation

varies according to altitude and longitude. Other radionuclides are human-made, such as those derived from fission reaction when producing energy. Assessing total exposure to ionizing radiation would therefore involve assessing radiation produced by a number of sources.

A radionuclide is characterized firstly by its chemical properties, secondly by the physical or physicochemical form in which it is found, and thirdly by its behaviour in biological media (chiefly its metabolic properties). It also has particular nuclear characteristics, namely, its disintegration rate (represented by its radioactive half-life corresponding to the time necessary for the disintegration of half the atoms present) and the nature and energy of the emitted radiation.

The human body, its tissues or organs can be exposed to radionuclides externally or internally. **External exposure** can occur when radionuclides are present in media outside the body and enter the body from the external media (this definition is similar to that used for exposure in Chapter 2 and elsewhere in this book, but is consistent with common usage when measuring radiation exposure), for example, when radiation emanates directly from a source such as a nuclear plant. This kind of exposure is almost exclusively occupational. External exposure also arises through use of diagnostic or therapeutic X-rays (mainly X-rays and γ-rays) and items such as luminous watches. X-rays and γ-rays can reach tissue some distance from the point of emission.

Internal exposure — involving penetrating γ-rays or much less penetrating α- and ß-particles — occurs when radionuclides are absorbed through inhalation or ingestion. (This definition is similar to that used elsewhere in the book for internal or absorbed dose, but again it is consistent with common usage when measuring radiation exposure.) For example, internal exposure can occur following inhalation of dust contaminated with adsorbed radon daughters. (Radon daughters are decay products of radon, which can themselves decay, releasing α-particles as they do so). After uptake, the radionuclides may remain in the organ where they were absorbed (e.g. contaminated dust inhaled in the lungs) or become distributed throughout the body.

Distinguishing between the two types of exposure is essential, since different measurement methods apply to each. Exposure that has resulted in tissue damage is assessed by measuring the energy absorbed by the tissue, taking into account the type of radiation. This is called "dose equivalent" and was formerly expressed in "rem". The joule per kilogram replaced this term (1 rem = 10^{-2}J/kg) in 1975; subsequently, the sievert (1 rem = 10m Sv) has become the term most commonly used. For external exposure, well developed and sensitive equipment, such as ionization chambers, scintillation counters, and thermoluminescent or photoluminescent dosimeters, can be used to measure the maximum dose equivalent received by the organism, directly from the radiation source. Alternatively, subjects can be equipped with direct-reading personal samplers.

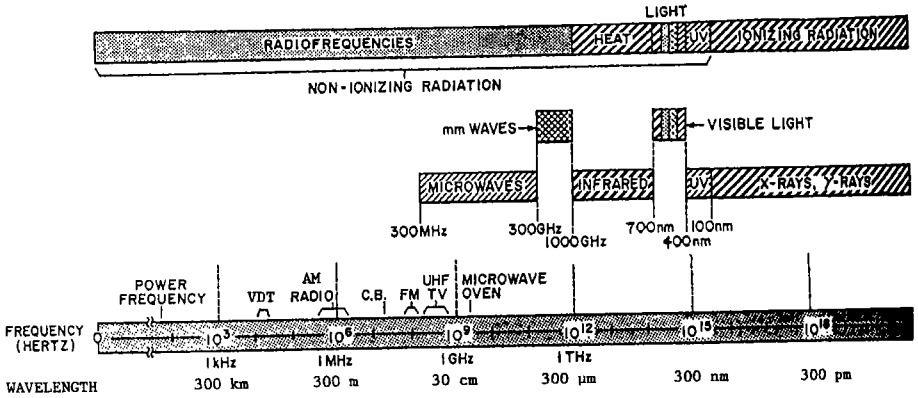
Dose equivalents cannot, however, be measured directly for internal exposure, since the source of exposure is inside the body. Instead, exposure is assessed by evaluating the radioactivity of radioactive materials, such as air, drinking-water and foodstuffs, that have been absorbed into the body. Many highly sensitive and accurate methods are available for calculating levels of such radioactivity. An example is an α-particle detector used to measure radon gas monitor or the concentration of radon gas in indoor air (or the α-particle radiation generated when the radon decays). The concentration of radon in air is used as an indicator of the internal exposure from inhaled air.

Non-ionizing radiation

Non-ionizing radiation refers to all radiation in the electromagnetic spectrum exclusive of the ionizing range (Shore, 1998). (See Figure 4.6.) It includes the various forms of light waves, microwaves and radio waves and is part of the natural atmospheric background radiation to which all life is exposed to a varying degree. Additionally, owing to technological advances, electronic sources have added greatly to environmental levels of non-ionizing radiation. The health significance of such exposures depends on the physical characteristics of the radiation, the conditions and duration of exposure, and the physical characteristics of the persons at risk.

Light radiation includes the ultraviolet, visible and infrared wavelengths of the electromagnetic spectrum. All are found in various proportions in sunlight. Various processes may also emit these wavelengths; in laser devices, the emissions are coherent monochromatic beams of light.

Figure 4.6. Electromagnetic spectrum in terms of frequency, wavelength & uses



Source: Shore, 1998; adapted from Petersen, 1983.

Exposure to non-ionizing radiation is measured as radiant energy. However, assessment of exposure intensities and durations is also required in order to account for temporal pattern and variations in exposure. These can be estimated by combining meteorological data with details (collected via interview) of personal habits concerning sunbathing, use of solariums, gardening, etc. Susceptibility to the effects of ultraviolet light — which is greater in light-skinned races than in dark-skinned races — must also be taken into account. An example of the use of the exposure approach in eye cataract studies is given in Box 4.4. (WHO has published a review of ultraviolet radiation that deals with many of these issues (WHO, 1994b).)

Electronic devices are the principal sources of microwave (MW) and radiofrequency (RF) radiation. Electromagnetic radiation occurs worldwide through use of radar, heating techniques, telecommunications, and broadcasting systems. Exposure to MW/RF fields is generally assessed by measuring average power density under specified conditions. For some sources, such as radar, peak power density may also have to be measured. Dosimetry is complex and international standardization of measurement techniques remains to be accomplished.

Box 4.4. Use of exposure approach in studies of eye cataracts due to solar radiation

To explore the expected range of dose values, a programme of studies was conducted at the US Army Environmental Hygiene Agency (US EHA) during 1994, based upon earlier experience gained in evaluation of human exposure to solar UVR during 1972–1994. To this end, ambient UVR measurements along with sky luminance and lid-opening measurements were made. Further experiments employed a mannequin to simulate the ocular geometry in sunlight for two scenarios: i.e. with and without sunglasses. The lid opening of individuals was measured under different outdoor conditions. An equation was then derived to estimate actual ocular UVR exposure in sunlight as a function of sky luminance, ground reflection and sky conditions. A series of studies was conducted on each of the following topics: environmental ultraviolet measurement; lid opening; and mannequin-based dosimetry.

Source: Sliney, 1995

Exposure assessment of physical factors requires an understanding of the factors' intrinsic properties and their behaviour in the environmental media. For example, noise or sound is energy waves transmitted through air. There is no sound in a vacuum. On the other hand, x-rays and ultra-violet light, for example, are transmitted across a vacuum. As another example, the α-radiation from the decay of radon in ambient air causes essentially no risk because the α-particles cannot penetrate the outer layer of the skin. However, the radon decay products can adsorb on to dust particles and be inhaled into the lungs, causing lung cancer from the α-radiation directly to the tissue of the airways. Therefore, the amount of dust particles in the air can affect the ultimate α-radiation dose from radon in air.

4.3 Exposure measurement techniques

Measuring exposures accurately and precisely is clearly of great importance when seeking to establish associations with health outcomes. Yet measuring exactly how much of an exposure reaches the target human organ is often impossible. Biological monitoring techniques can provide good estimates of dose but they are often not practical, or available, when assessing exposure to environmental pollutants. In some cases, individual exposure levels can be measured by using personal monitors. However, obtaining this information for a sufficient number of individuals is rarely feasible since the expense involved would be considerable. Moreover, obtaining the cooperation of large numbers of subjects for long-term personal monitoring is often difficult. In practice, therefore, exposure variables are generally approximations or **surrogates** of the actual exposure. Unfortunately, the accuracy and precision with which true exposure is approximated may vary widely from one surrogate exposure variable to the next. A hierarchy of exposure data suggested for measuring exposure to a fixed point environmental source is therefore presented in Table 4.7. This example of a hierarchy is not shown, though, to suggest that it is always preferable to measure the "highest" (i.e. most accurate and precise) variable in the hierarchy. The choice of exposure variable will depend on a number of factors including cost and feasibility. A broader framework for considering potential hazards is the DPSEA framework described in Section 1.3 and illustrated in Figure 1.4, pp. 11 (Kjellström & Corvalán, 1995).

Table 4.7. Hierarchy of exposure data and surrogates for fixed source contaminant

Types of data	Approximation to actual exposure
1. Quantified personal measurement	Best
2. Quantified area measurements in the vicinity of the residence or sites of activity	
3. Quantified surrogates of exposure (e.g. estimates of drinking-water use)	
4. Distance from site and duration of residence	
5. Distance or duration of residence	
6. Residence or employment in a geographic area in reasonable proximity to site where exposure can be assumed	
7. Residence or employment in a defined geographical area (e.g. a county) of the site	Poorest

Source: adapted from NRC, 1991 a.

No universal guidelines can be given on how to decide which measurement techniques to apply in any given study. The final choice will be the result of a process involving consideration of:

- validity;
- cost;
- experience;
- availability of exposure measurement equipment;
- laboratory facilities.

The availability of resources, in particular, will vary from country to country and often even within countries. That said, prior assessment of the validity and reproducibility of available exposure measurement techniques must be undertaken to decide whether a study has the potential to produce a valid result. Generally during the planning stage of a study, some evaluation of the quality and availability of exposure data is necessary in order to decide whether a study is indeed feasible.

Questionnaire data

Questionnaires are an important source of information for exposure assessment. The principal role of questionnaires is to obtain information to guide the exposure measurement, and to link individual subjects to environmental measurements when comprehensive personal environmental sampling is not feasible. Individual subjects may be able to report on potential exposures in their household or community. This information can then guide the epidemiologist when undertaking the exposure measurements. Subjects also may report on their symptoms in a questionnaire, which might guide the environmental sampling. For example, subjects may report that their symptoms worsen in a particular area of a building or when in a park near a waste discharge drain. Environmental measurements would then be taken in those locations.

An important use of questionnaires or diaries is that of determining the proportion of time that a person spends indoors at home, or at school or workplace, or in other locations. This information can be used to calculate the integrated exposure if environmental measurements are obtained in each of the locations (see the section on micro-environments, pp.92).

Questionnaires may have to be used as a primary source of information about exposure if direct personal or environmental measurement would be too expensive or complex. For example, as described on pp. 91, a questionnaire may be the most feasible method for gathering information on an individual's food and drink consumption. Questionnaires are also used to collect information that can serve as surrogates of exposure. Studies have shown, for example, that responses to questions about whether a household has a gas or electric stove correlate with measurements of indoor air concentrations of nitrogen dioxide (NO_2) (see Table 4.8.)

Table 4.8. Relationship between use of gas versus electric cooking appliance and weekly average NO_2 concentration

	Average NO_2 concentration in $\mu g/m^3$ (s.d. in parentheses)	
Cooking appliance	Kitchen	Other room
Gas stove	92.3 (43.4)	46.8 (30.5)
Electric stove	15.6 (9.4)	12.8 (6.0)

Source: adapted from Palmes et al., 1977.

Similarly, questions about whether a parent (or caregiver) smokes cigarettes have been found to be predictive of a child's environmental tobacco smoke exposure, as indicated by urinary or salivary cotinine concentrations. Yet another example concerns the reporting of odours on questionnaires which can be an indicator of whether exposure concentrations are above or below the odour threshold. (Experimental studies have found that subjects tend to report odours of chemical substances at fairly consistent concentrations.) Another common example of an exposure surrogate obtained from a questionnaire is that of the location and duration of a person's residence in a community.

Finally, questionnaires can be used to obtain information from subjects concerning factors that may modify the potential effect of an exposure on a health outcome. For example, information on personal habits such as smoking or alcohol consumption is important in many environmental epidemiology studies. As illustrated in Table 4.1 and Table 4.2, work and activity patterns, which can be reported on a questionnaire, influence exposure and dose.

Questionnaires are typically administered at one point in time or at fairly long time intervals during the course of a study so there may be problems with individuals' recalling detailed information. An related approach to address this issue is to have the person keep a **diary** of the relevant information. A diary is a generally small form which the individual keeps with him or herself to record information on a pre-defined time interval. The time interval can be daily or even as short as every 15-minutes. Use of a diary is particularly relevant to record time and activity patterns, consumption of foods and drinks, and for health outcomes such as symptoms.

Biological markers

The role of biological markers in epidemiological studies is discussed in Chapter 2, pp. 25. Since this chapter is on exposure assessment, the focus here is on markers of internal and biologically effective dose (which still are commonly referred to as biomarkers of exposure) (Griffith et al., 1989). Such markers include measurement of pollutants and their metabolites in body tissues and fluids, and reflect concentrations of pollutants that have entered the body. Specific markers may reflect instead physiological effects or organ system damage — but can also be interpreted as measures of exposure. For example, cholinesterase activity level in red blood cells and plasma are used as a biological marker of exposure to organophosphate pesticides, which alter cholinesterase activity irreversibly by binding with the cholinesterase enzyme.

In effect, biological markers integrate exposure for various pathways. This means that measuring exposure for each of the different environmental pathways is unnecessary. This integration can be a disadvantage though in those environmental epidemiology studies that aim to evaluate the importance of different environmental exposure routes. In such cases, the relationship between the marker(s) of dose and exposure through various environmental pathways may need to be worked out or determined (Brunekreef et al., 1983).

Measuring personal exposure

A variety of sampling devices have been developed to measure exposure to air contaminants. These devices generally fall into two categories — active and passive sampling devices. An active sampling device consists of a filter or other sampling media attached to a small air pump which actively draws air at a known rate across the sampling media. The concentration of air contaminant is calculated by dividing the amount collected in the sampler by the volume of air sampled by the pump. In a passive sampling device, there is no air pump and so the air is sampled passively due to natural air movement and diffusion of the contaminant through the air. Passive sampling devices are also commonly used to measure radiation exposure. An example familiar to many people who have worked in a medical facility is the radiation badge worn by radiology department technicians. The badge is worn by the technician for a period of time and then analysed to determine the amount of radiation to which the technician has been exposed.

Many instruments now being used to measure personal exposure in the general environment were originally developed for use in occupational settings. But exposure to contaminants is generally higher in the occupational environment than in the general environment. Therefore, techniques developed to measure occupational exposure are not necessarily appropriate for measuring exposure in the general environment. Factors that should be taken into account when considering personal monitoring in the general environment include:

- the detection limit of the method;
- the complexity of contaminant mixtures (as these may contain substances that interfere with measurement of the substance of interest);
- the length of time for which battery-operated equipment can run before needing to be recharged;
- the possible disruption of normal daily activities caused by wearing or operation of the instrument.

If any of these factors present significant difficulty, area monitoring rather than personal monitoring may have to be undertaken instead. See Table 4.9 for a comparison of the strengths and weaknesses of personal and area monitoring.

Table 4.9. Strengths and weaknesses of personal and area monitoring

Type of monitoring	Strengths	Weaknesses
Area	Sophisticated equipment; large sample volumes; can be used continuously	Does not necessarily represent personal exposure
Personal	Represents personal exposure	Volume restrictions; weight of equipment; small sample volumes; cannot be used continuously

Food and Water

Techniques can be used to measure personal exposure to contaminants that occurs via routes other than air. An accurate, although intensive, means of measuring exposure to contaminants in food is that of the **duplicate diet** study in which exact duplicates of all foods and beverages consumed by subjects are collected and analysed for contaminants. Practical constraints limit application of these techniques to the collection of meals for between one and a few weeks, from a limited number of subjects. The exercise can be repeated, but the demands on individual participants are quite high. Ideally, the study subjects should be supervised.

Indirect assessment of intake can be undertaken by means of "total diet" or "market basket" studies, selective studies on individual foodstuffs, or "habit surveys" (nutrition table method).

In market basket studies, food samples of the principal diet constituents, such as cereals, meats, root vegetables, etc., are prepared, based on national consumption data. After normal preparation and cooking they are analysed. The mean concentration of toxic elements in each constituent is then measured. An average daily intake can then be calculated for each item and for the diet as a whole if data on dietary habits is available. Such studies are repeated for different seasons and in different regions to reflect local variations in the diet. These studies are particularly valuable if elements (e.g. lead, cadmium) are widely distributed among all major food items, or, as with mercury and arsenic, if bioconcentration occurs almost exclusively in fish and shellfish.

In studies of **individual foodstuffs**, concentrations of contaminants are measured in representative samples of staple foods. The most commonly-occurring levels, together with food consumption data, can be used to calculate average daily intakes. This approach is particularly appropriate if intake consists primarily of one or two items or types of food, or if food monitoring programmes have established an average concentration of pollutant in a particular commodity (e.g. DDT in cereals).

In **habit surveys**, a sample is selected from a population of particular interest, such as the young or the elderly, who may be more susceptible to the effects of a biological or toxic chemical agent) to obtain information about the most extreme food consumption habits of these population subgroups. In the United Kingdom, such an approach, using interviews, was adopted to determine the consumption habits of people in the upper 5% of consumption rates). A habit survey is thus a reasonably accurate reflection of the consumption levels of the most extreme and susceptible consumers. This method can help identify consumers who experience unacceptable exposure and/or increased metabolic susceptibility to toxic elements because of their food consumption patterns.

Micro-environmental measurements

Micro-environmental measurements refer, for example, to measurements of air pollution taken in homes or means of transportation, measurements of drinking-water contaminants in tap water, and measurements of contaminants in foodstuffs bought by individual households.

Nitrogen dioxide (NO_2) exposure has often been measured for epidemiological studies by means of micro-environmental sampling and, therefore, can serve as an example of this technique. In the mid-1970s, passive sampling techniques that relied on inexpensive and small instruments were developed to measure NO_2 in air (Palmes et al., 1976). Several epidemiological studies on the respiratory health effects of indoor NO_2 pollution caused by use of unvented gas appliances have used such samplers to measure indoor NO_2, rather than predicting it from the presence and usage patterns of the appliances (see Table 4.8; see also, for example, Dijkstra et al., 1990; Neas et al., 1991). Additionally, several investigators have examined how such measurements correlate with personal exposure to NO_2, how such daily or weekly averages relate to short-term peak exposures due to intermittent use of gas appliances, and the degree to which study precision increases if actual measurements of NO_2 rather than surrogate exposure variables are used (Noy et al. 1990; Schwab et al., 1993; Xue et al., 1993).

Time-activity patterns

To assess human exposure to environmental contaminants accurately using measurements of micro-environments, an estimate of where and how subjects spend their time is required. "Time-activity patterns" can be established in a number of ways. As discussed above, subjects can be requested to complete questionnaires or equipped with diaries and/or data-logging instruments, and requested to register where and how they spend each 15-minute period. Commonly-occurring micro-environments (such as bedroom, living office, or car) can be pre-coded. Others can be entered as needed. Ceiling distance sensors are a recent innovation. A gadget worn by an individual, such a sensor can differentiate between in-transit micro-environments (the "ceiling" is less than 1.2 m away from the sensor), indoor microenvironments (the ceiling is between 1.2 and 3.35 m distant from the sensor) and outdoor environments (the ceiling is more than 3.35 m away from the sensor) (Moschandreas et al., 1993). Subjects are sometimes also requested to assess their level of physical activity so that the volumes of air that they inhale per minute can be estimated.

An example of a time-activity study can be found in Schwab et al. (1990). Personal exposure to NO_2 was investigated in a population of several hundred subjects living in Los Angeles. The subjects were requested to maintain a time-activity diary on the days when personal exposure

measurements were taken. The time spent in each of the seven micro-environments is shown in Table 4.10. Micro-environments were defined not only for location, but also for activity ("kitchen with gas range on", "kitchen with gas range off"). The data in the table show that most time (more than 16 hours on average per day) was spent in the home. Subjects also spent nearly three hours per day outdoors; more than half that time was spent near major roads.

Table 4.10. Time (in hours) spend in each micro-environment during study of NO_2 exposure

	Total sample (n = 1283)		Persons reporting time in microenvironment
Micro-environment	Mean (± std. dev.)	Percent reporting participation	Mean (± std. dev.)
Kitchen — range on	0.49 (±0.86)	49.2	1.00 (±1.00)
Kitchen — range off	0.75 (±1.35)	52.2	1.36 (±1.58)
Inside home — other	14.89 (±4.28)	99.5	14.97 (±4.14)
Inside work or school	3.36 (±3.85)	50.8	6.62 (±2.75)
Inside — other	1.69 (±2.68)	57.3	2.95 (±2.97)
Outside — major roads	1.71 (±1.86)	79.4	2.15 (±1.85)
Outside — other	1.16 (±1.94)	52.9	2.20 (±2.19)

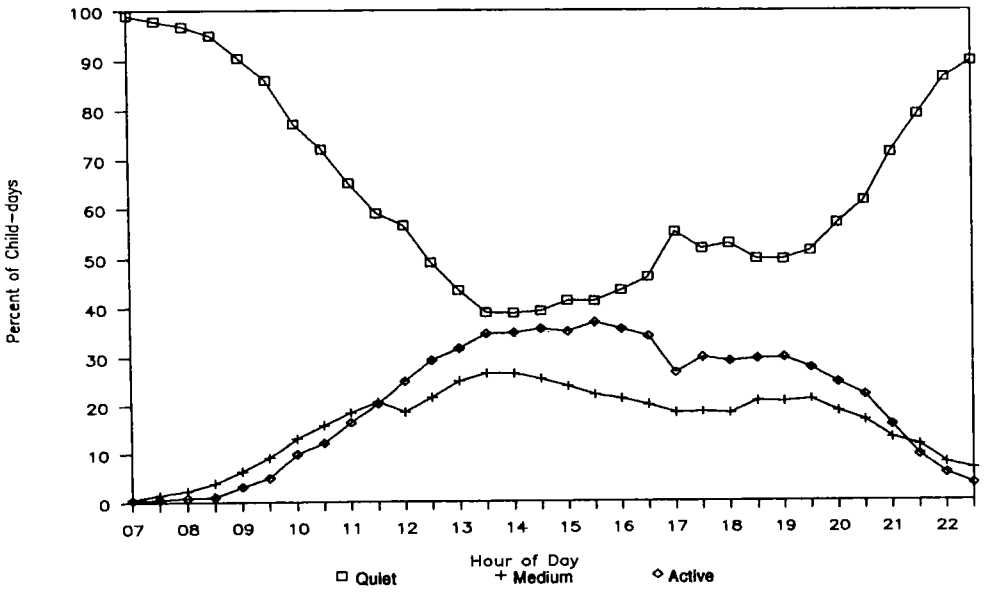
Source: Adapted from: Schwab et al. (1990).

An example of the "self-reporting" by subjects of their activity levels is provided by the Schwab et al. (1991) study of a group of 9–11 year old children. Three different levels of exertion were defined:

> "*Quiet*; you were fairly still and quiet, includes reading, coloring, sleeping, watching TV, or computer games.
>
> *Medium*; you were busy moving around, but not too fast. Included playing inside, doing chores, walking to your friend's house.
>
> *Active*; you were using a lot of energy and breathing hard. Includes sports and games with friends, running, perhaps a lot of laughing and talking."

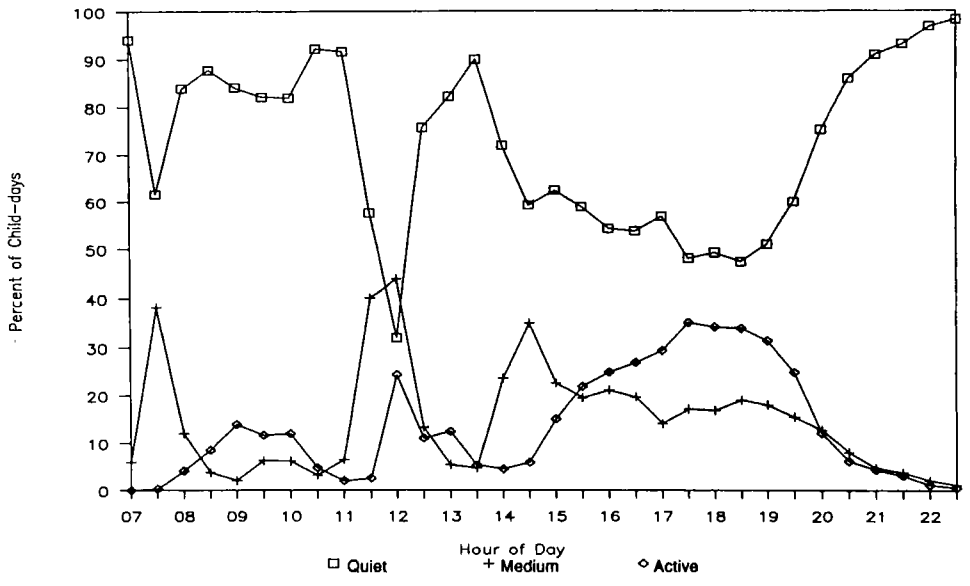
The amount of time spent in various micro-environments at the three predefined activity levels was determined. The data showed that, on average, the children considered themselves active for about three and a half hours each day. Much of the active time was spent outdoors, whereas most of the quiet time was spent indoors. A marked difference was observed between activity levels, depending on the time of day, and type of day (schoolday, non-schoolday), as seen in Figure 4.7 (non-school days) and Figure 4.8 (schooldays).

Figure 4.7. Activity levels according to dime of day and type of day – schoolday



Source: Schwab, 1991.

Figure 4.8. Activity levels according to dime of day and type of day – non-schoolday



Source: Schwab, 1991.

Time-activity data can be combined with micro-environmental measurements to estimate personal exposure integrated over a certain period. Noy et al. (1990) compared the personal exposure to NO_2 of schoolchildren, with exposure estimated in three different ways:

- using information on how time was spent, and NO_2 concentrations measured in three different rooms (kitchen, living room, bedroom) in the home;
- using the average NO_2 concentration of the three rooms;
- using the living-room NO_2 concentration as surrogate.

The correlations between personal exposure and the three different exposure estimates were comparable. This showed that for this population, and for this pollutant, exposure could be estimated reliably on the basis of stationary measurements taken for one room of the home. Taking measurements in other rooms was not necessary. Neither was collecting information on individual time-activity patterns. In this case, time-activity patterns did not vary significantly since the subjects spent much of their day together in the same place (at school). The results could not be extrapolated to other populations with more heterogeneous time-activity patterns.

Time-activity patterns are most commonly established for studies of the health effects of air pollution, but can also be used for studies of physical factors such as solar UV radiation, electrical and magnetic fields, and pollutants in soil. If, for example, the direct exposure of children to soil pollution is being assessed, information is required on how much time children spend at contaminated sites.

Time-activity patterns are not relevant for measuring exposure to contaminated water and food; however, the same concepts are used for the diet studies described. Specifically, the epidemiologist uses questionnaires or diaries to determine the specific type and amount of food consumed (cf. time-activity diaries) and then estimates or measures the contaminant concentrations in samples of each of the food items (cf. obtaining environmental measurements for each micro-environment). The information from these two sources is then combined to calculate the integrated exposure.

Macro-environmental measurements

Macro-environmental measurement refers to environmental monitoring on a large spatial scale, such as a neighbourhood or a region. Traditionally, studies of the health effects of outdoor air pollution linked the health status of communities to air pollution concentrations measured at one or several monitoring sites. In terms of an individual subject's exposure, measurements at a macro-environmental level are clearly less precise than personal or micro-environmental measurements. But macro-environmental measurements may be the only environmental monitoring data available. Furthermore, macro-environmental monitoring may be suitable in situations where contaminants are dispersed relatively consistently. For example, measuring fluoride and arsenic concentrations in a municipal water supply, as an indicator of potential drinking-water exposure to these contaminants, may be appropriate because the concentrations anywhere in the water supply system are likely to be similar.

Surrogate exposure variables

If contaminants themselves cannot be measured, measurement of variables that correlate with contaminant concentrations ("surrogate exposure variables") may provide an alternative. In actual practice, the vast majority of environmental epidemiology studies must rely to some extent on the use of surrogate exposure variables since the cost and complexity of obtaining biological or environmental measurements for large numbers of subjects are generally prohibitive. Table 4.11 presents examples of typical surrogate exposure variables.

Table 4.11. Examples of surrogate exposure variables

Outdoor air pollution	"urban" versus "rural" density of motor cars
Environmental tobacco smoke	live with a smoker proportion of population who are smokers
Indoor radon	building construction (building material, amount of airflow through building)
Magnetic fields	wiring configuration
Water contamination	well or public supply source
Diet	main type of domestic agriculture
Point source (e.g. factory)	distance from source

Distance from a contaminant source

Contaminant concentrations tend to decrease rapidly with distance due to dispersion in environmental media. The distance from a contaminant source is therefore often used as a surrogate measure of exposure in community-based studies of single or fixed-source contaminants, such as an industrial facility or hazardous waste site. The source does not have to be a point source, however. For example, studies of water contamination have used distance from a river as a surrogate exposure variable, and studies of lead air pollution have measured distance from major roadways as an exposure variable in countries where petrol that contains lead is still used.

Another example concerns analysis of 107 cases of Legionnaires' disease detected at different times between 1978 and 1986 in Glasgow, UK. The relative risk of developing this disease appeared to be related to the distance of people's homes from a cooling tower, which was the apparent source of the microorganisms responsible for the disease (see Table 4.12). Another example of this type is shown in Table 4.13. It illustrates that the percentage of households with young children with trachoma is dependent on the distance to the nearest water source.

Table 4.12. Relationship between distance from a cooling tower and relative risk of Legionnaires' disease in Glasgow

Distance from cooling tower	Relative risk (95% CI)
< 250 m	3.89 (1.92–7.70)
250–500 m	3.00 (1.79–5.02)
500–750 m	1.19 (0.62–2.22)
750–1000 m	1.11 (0.57–2.11)
> 1000 m	1.00 (reference distance group)

Source: adapted from Bhopal et al., 1991.

Sources of contaminants in the home

As discussed in the section on questionnaire data, several studies have now been published that have used measurement of indoor NO_2 concentrations to assess exposure. These studies were initiated following early indications of a risk associated with gas cooking (Melia et al., 1977) and publication of a report demonstrating a clear relationship between the use of unvented gas appliances and indoor NO_2 concentrations (Palmes et al., 1977). Other sources of exposure to contaminants in or close to the home include:

- open fireplaces (source of wood smoke)
- unvented kerosene heaters (source of various combustion products)
- damp walls (source of moulds, or an indicator of the potential of the home to sustain populations of dust mites)
- furniture constructed from particle board (source of formaldehyde)
- lead water pipes (source of lead in tap water)
- high voltage transmission lines (source of electric and magnetic fields)
- homes located in busy streets (as an indicator for exposure to traffic-generated pollutants).

Table 4.13. Percentage of households with children aged 1–7 years with trachoma, according to the time (distance) to the nearest water source

Time to water source	Trachoma status of household (percent) *		
	no children affected	some children affected	all children affected
< 30 minutes (389)**	38	25	37
0.5–2 hours (844)	27	24	49
> 2 hours (705)	27	21	50

* $p < 0.001$

** Figures in parentheses are the number of households investigated

Source: WHO, 1991a.

4.4 Exposure modelling techniques

Estimation through modeling has been developed as an alternative to measurement of personal exposure. Modeling involves estimation of personal exposure using mathematical formulae based on data and knowledge of variables other than actual direct personal measurements. Conceptually, exposure models are used to estimate exposure whenever an investigation relies on surrogate variables, because direct exposure is not being measured. It is important to recognize that these approaches conceptually involve modeling so that assumptions implicit in the models about the relationship between the surrogate variable and actual exposures are made explicit.

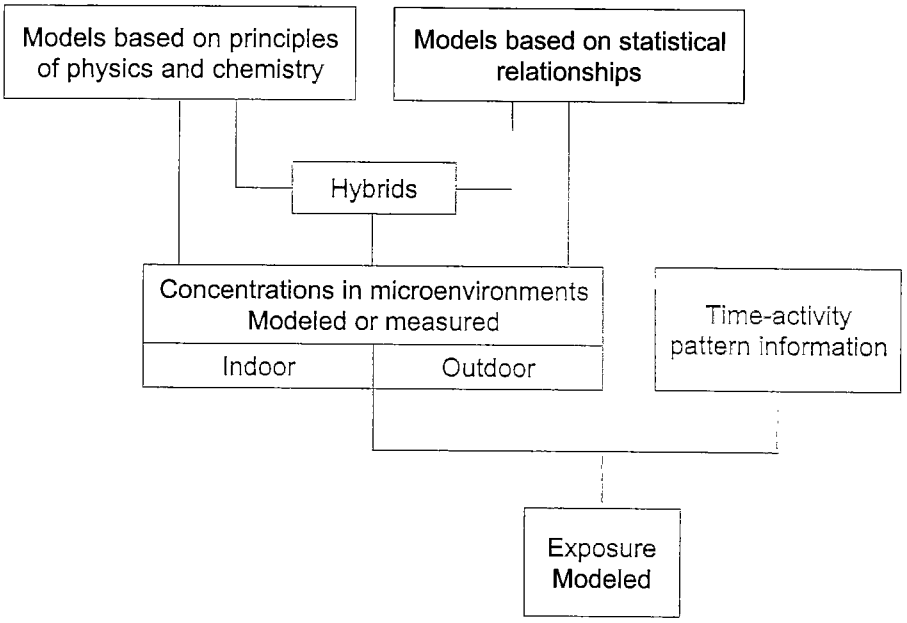
The application of exposure models may be straightforward; at the other extreme, it may be highly complex. Referring to examples presented earlier in this chapter, a simple exposure model could be constructed to classify an individual's indoor exposure to NO_2 on the basis of the single surrogate variable of having a gas versus an electric stove. The **integrated exposure model** is more complex but nevertheless common. It models exposure to an airborne contaminant by combining the time-activity variables and the micro-environmental measurement data (see Fig. 4.3). This approach is considered as an example of modelling since personal exposure is not measured directly and continuously throughout the study. The measurements for the microclimates are combined in the integrated exposure model. Another example concerns the calculation of total exposure to a foodborne contaminant using dietary information and food content data. Models can also incorporate knowledge about toxicokinetics (i.e. rates of absorption, metabolism, distribution, storage, and excretion of an xenobiotic) to estimate internal dose and biologically effective dose using data on environmental concentrations. These latter models are generally referred to as toxicokinetic or pharmacokinetic models.

Two general classes of model exist for estimating exposure: "deterministic" (i.e. physical) and "stochastic" (i.e. statistical). **Deterministic models** describe the relationships between variables mathematically, on the basis of knowledge of the physical, chemical and/or biological mechanisms governing these relationships. An example of a deterministic model would be the calculation of target organ dose of cadmium using data on environmental concentrations in various environmental media, combined with knowledge about anticipated contact and absorption rates, as well as toxicokinetic information about cadmium following absorption.

Stochastic models express the statistical relationships between a determinant of exposure, and exposure itself, inferred from data collected in situations similar to that for which exposure is being assessed. Stochastic models do not necessarily require fundamental knowledge of the underlying physical, chemical and/or biological relationships between the variables. Estimation of exposure to NO_2 in indoor air, based on data about the type of cooking appliance (gas versus electric) and empirical knowledge about NO_2 levels reported in earlier studies, is an example of a stochastic model. (In this instance, knowledge of the physical and chemical properties of NO_2 and indoor air pollution is unnecessary.) A third class of model is sometimes described as **hybrid** or **physical-stochastical**. These combine aspects of both the deterministic and stochastic models. Figure 4.9 is a schematic diagram of the models used in exposure assessment.

Sexton and Ryan (1988) compared these three types of model with reference to estimation of pollution exposure. Their overview of the different characteristics of each model is presented in Table 4.14.

Figure 4.9 Schematic diagram of models used in exposure assessment



Source: NRC, 1991b.

Table 4.14. Comparison of different approaches to exposure modeling

	Model type		
Parameter	Deterministic (physical and biological)	Stochastic (statistical or empirical)	Hybrid
Method of formulation	Physical laws	Hypothesis testing (empirical statistics)	Physical laws and statistics
Required input	Knowledge of important parameters and their values in the system to be modelled	Data on human exposure	Knowledge of important parameters and their distributions in the systems to be modelled
Advantages	True model developed from *a priori* considerations	Makes use of real data in the model-building process	Model developed from *a priori* considerations; stochastic part allows uncertainty to contribute
Disadvantages	Includes investigator's biases; must be validated	Requires data on hand for model building; extrapolation beyond data base is difficult	Requires much knowledge of systems; must be validated

Source: adapted from Sexton & Ryan, 1988.

Deterministic models

The Indoor Air Quality model, recommended by the US National Academy of Sciences, is a deterministic model for predicting indoor air pollution concentrations for individual homes (NRC, 1981). It uses measurements of or assumptions about the values of the factors governing indoor air pollution concentrations using the following parameters: the concentration of a pollutant in ambient air, ventilation rate, contaminant decay rate and production rate of contaminant indoors. The structure of the model is as follows:

$$V\frac{dC_i}{dt} = P - E - Q*(C_i - C_o)$$

where:

V	=	volume of indoor space (m^3)
C_i	=	indoor pollution concentration ($\mu g/m^3$)
t	=	time (h)
P	=	production rate of pollutant in indoor space ($\mu g/h$)
E	=	elimination rate of pollutant from indoor space through chemical and physical processes ($\mu g/h$)
Q	=	air exchange rate with exterior (m^3/h)
C_o	=	outdoor pollution concentration ($\mu g/m^3$).

Data or knowledge about each of these parameters would be necessary in order to calculate the indoor pollution concentration for any point in time.

Models have also been constructed for estimating the ambient air concentration of a pollutant at a given point in space and time, based on knowledge of the dispersion and chemical reactivity of pollutants in the atmosphere. (See also NRC, 1991b, for further discussion.) If they are combined, and if they incorporate information on time-activity patterns, indoor and outdoor air quality models can be used to estimate personal exposure to air pollutants. A simple model of personal exposure is the time-weighted average of the indoor and outdoor concentrations, as follows:

$$C_p = C_o*\frac{t_o}{t_o+t_i} + C_i*\frac{t_i}{t_o+t_i} = [C_o*t_o + C_i*t_i]/(t_o+t_i)$$

where:

C_p	=	personal exposure
C_o	=	outdoor air concentration
t_o	=	time spent outdoors
C_i	=	(home) indoor air concentration
t_i	=	time spent (at home) indoors.

If the periods of time spent in micro-environments not covered by the model are short, such a model will yield valid exposure estimates (all other factors being equal). But if a significant proportion of time is spent in a micro-environment in which the pollutant concentration is very different from either the outdoor or the indoor concentration, important deviations may occur.

As a more general approach, time-activity data can be used with measurements taken in multiple micro-environments to estimate personal exposure integrated over time. The estimated, integrated personal exposure for subject "i" is the sum of the products of time spent in "j" separate micro-environments and the pollutant concentration in those microenvironments, or:

$$E_{ij} = \sum_{j=1}^{J} (c_{ij} * t_{ij}) \qquad \mu g/m^3 * h$$

Table 4.15 shows an example of the integrated personal exposure model for an individual who passes through a number of micro-environments during a 24-hour period, encountering different CO concentrations. These add up to a 24-hour cumulative exposure. The cumulative CO exposure in this example is 93 mg/m^3-h, or a time-weighted 24-hour average equal to 3.88 mg/m^3 (i.e. 93/24 h).

Table 4.15. Example of micro-environments occupied by person and corresponding carbon monoxide concentrations over a 24-hour period

Time	Micro-environment	Carbon monoxide concentration	Duration	E_{ij} (mg/m^3-h)
00h – 07h	Home	1	7 h	7
07h – 08h	In transit	17	1 h	17
08h – 12h	Office	4	4 h	16
12h – 13h	Restaurant	9	1 h	9
13h – 17h	Office	4	4 h	16
17h – 18h	In transit	18	1 h	18
18h – 20h	Shop	3	2 h	6
20h – 24h	Home	1	4 h	4
Total			24 h	93

Source: WHO, 1982.

The above examples are relatively simple because they account for exposure to one pollutant occurring primarily via one pathway. In principle, the same approach is used to develop more complete models which take into account multiple exposure pathways and variables, such as activity and breathing pattern, that can affect absorption. The limitation in using these models is that knowledge about all the physical and biological parameters may be incomplete, and even if reasonably complete, the needed environmental concentration data may not be available to feed into the models.

Recently, epidemiologists have recognized the important distinction between potential exposure (or environmental concentration) and the biologically effective dose when trying to determine precisely the effect of a toxic agent on disease. More attention has therefore been given to understanding biological factors that influence the biologically effective dose. Toxicokinetic or pharmaco-kinetic models are being developed to estimate the biologically effective dose using environmental measurement data. These pharmaco-kinetic models complement exposure models constructed to estimate the complete environmental hazards pathway (see Figures 4.1 and 4.2).

Stochastic models

Stochastic models differ from deterministic models in that a statistical or empirical approach is adopted to model the relationship between the available data and estimated exposure. No assumptions or knowledge of physical or biological parameters are used to create the models. But stochastic models do require data on exposure and on the factors believed to have influenced exposure. The relationship between exposure and the potential determinants of exposure are then explored using statistical techniques. Because this relationship is determined primarily by the data, and not by assumptions about mechanisms governing the relationship between exposure and determinants of exposure, care must be taken when extrapolating stochastic models beyond the data and populations on which they were originally based.

For example, Sexton et al. (1983) described the development of a stochastic model for indoor exposure to NO_2. Weekly average NO_2 concentrations were measured in Portage, Wisconsin, USA, in 25 homes that used electricity for cooking, and in 112 homes that used gas for cooking, on eight different occasions over a one-year period. They derived the following empirical model:

$$C_i = a + 0.6C_o \ , \qquad \text{where } a = 0 \text{ when electricity is used for cooking, and } 45\ (\mu g/m^3) \text{ when gas is used}$$

C_o = outdoor NO_2 concentration
C_i = indoor NO_2 concentration.

The values for "a" and the "0.6" coefficient for C_o were derived from the data on the 137 homes. The model was based on only two readily available variables — i.e. the use of a gas stove and the outdoor NO_2 concentration. The model did not attempt to identify data on all of the parameters for the NAS Indoor Air Quality model described in the above section. For example, the relationship between the use of gas for cooking and exposure to NO_2 indoors may be influenced by appliance usage, room volume and ventilation, and the presence of other indoor sources of NO_2 such as unvented kerosine heaters and unvented gas-fired water heaters (Palmes et al., 1979; Fischer et al., 1986). Nevertheless, this empirical model was felt to provide an adequate estimate of indoor NO_2 exposure. The model was subsequently applied to estimate indoor exposure to NO_2 in six different cities in the US, using data on the outdoor NO_2 concentration, and on the percentage of homes using gas or electricity for cooking and the results were used in an environmental epidemiology study.

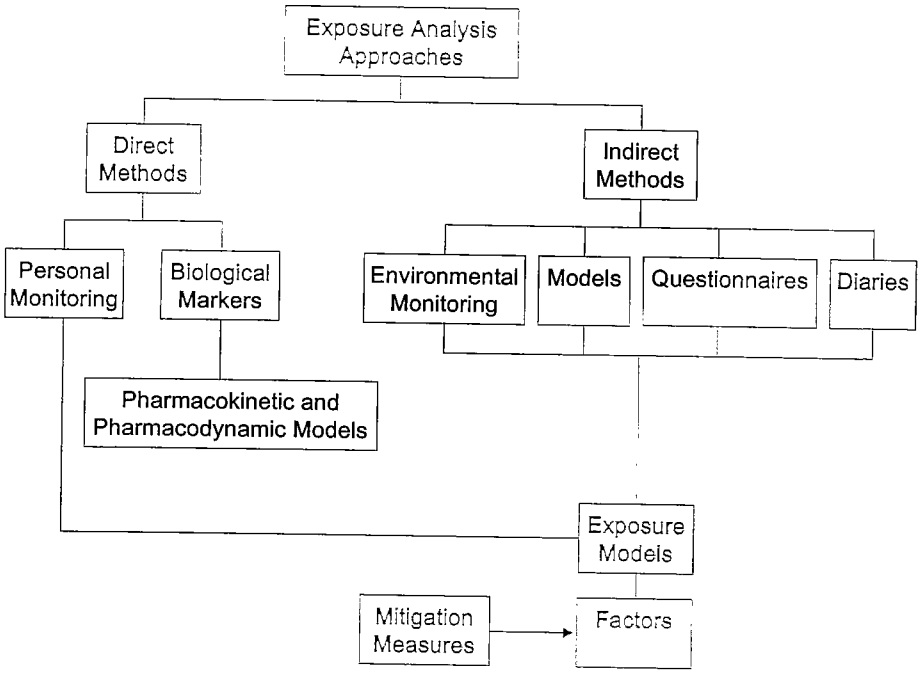
Given that epidemiologists should attempt to use as much knowledge as possible about the physical parameters governing exposure and as much data as is available about human exposures in the population, no clear distinction can be made between deterministic and stochastic models. Most exposure modeling would incorporate as much of the available data as possible and would therefore be hybrid models.

4.5 The exposure variable in an epidemiology study

Exposure assessment is the process of selecting appropriate exposure variable, measuring the exposures and then interpreting the variable for use in the epidemiological study. The exposure assessment should incorporate all of the information derived from the environmental

measurement and modelling techniques described above. Figure 4.10 illustrates how exposure to air contaminants would be analysed.

Figure 4.10. Approaches for analysis of air contaminants



Source: Sexton & Ryan, 1988.

Selection of the exposure variable

Choice of the appropriate exposure variable and an exposure assessment strategy is a critical step when designing an epidemiological study. Sometimes, several exposures are considered for separate epidemiological analysis or for a single analysis based on the principle of combined exposure. The choice of variables must be based on an understanding of the physical laws governing the transformation and dispersion of pollutants (or hazards) in environmental media. The potential for exposure is also affected by the physio-chemical or biological properties of the agent(s) and the pharmaco-kinetic and biological mechanisms governing absorption. Each of these factors should be taken into account when selecting exposure variables.

Even if technically possible and ethically feasible, measuring biomarkers of internal or biologically effective dose, or personal exposure, is usually much more expensive than using simpler exposure measures. For example, study subjects must be contacted individually so that blood samples can be collected, or requests made that they wear personal monitors or collect duplicate diets. If personal monitors are used, a large number may be needed to ensure sufficient statistical power, and large numbers of samples analysed in the laboratory. Yet funding for exposure assessment as well as for the other parts of a study is likely to be limited.

Armstrong (1993) has provided a general framework to help investigators decide which measures of exposure to use to obtain maximum statistical power. This approach would generally be used only by researchers with advanced training, but the concept is presented here to give an overview of the approach. The basic premise of Armstrong's considerations is that measuring true exposure is better, but more expensive, than measuring approximate exposure. However, if the correlation between the approximate and the true exposure variable is high, the loss of power entailed by using the approximate rather than the true exposure variable will be small. Accordingly, if the cost of measuring approximate exposure per study subject is clearly lower than the cost of measuring true exposure, a study using the approximate measure of exposure will be more efficient.

Alternative study designs can be compared by calculating the "Asymptotic Relative Efficiency" (ARE), defined as the ratio of the sample size necessary to achieve equal power (in each study design) to detect an association (Armstrong, 1993). If "r" is the correlation between the approximate and the true exposure, "C_i" the basic cost of including a subject in the study (on assessment of disease status, for instance), and "C_a" and "C_t" the cost of measuring approximate and true exposure, respectively, per subject, then the ARE can be expressed as:

$$ARE_{a/t} = r^2 [(C_i + C_t) / (C_i + C_a]$$

So if the correlation between the approximate and the true exposure variable is equal to 0.5 ($r^2 = 0.25$), use of the approximate variable will be more efficient than use of the true variable, provided that the total cost of including subjects to measure the approximate variable is more than four times (1/0.25) lower than the cost of including subjects to measure the true exposure variable. Such differences in cost of this or even a larger magnitude are likely to occur if the choice is between taking personal exposure measurements for all subjects and collecting information on sources, habits and/or occupations by questionnaire, and if the "fixed" cost of including study subjects (C_i) is not too high compared to the cost of taking exposure measurements *per se*. Box 4.5 illustrate these considerations.

A decision to use approximate exposure variables must be based on knowledge of the correlation between these variables and the true exposure. Frequently, this information has to be obtained in a pilot study. This adds to the cost of the total study. Armstrong (1993) also provides guidance for the relative allocation of budget between the pilot and main studies.

Individual versus group data

Almost all the approaches for measuring exposure in individual subjects are costly and time-consuming, and consequently cannot be applied easily to a large number of individuals as part of an epidemiological study. More commonly, therefore, exposure is assessed on the basis of measured exposure levels for a defined area within the environment. The measured levels are then used to assign an exposure score to all individuals living or working within the area. Such an approach necessarily ignores individual variations in exposure and results in some misclassification of exposure levels. These errors are likely to increase with increasing size of the individual area units. Therefore, the use of small units (e.g. at the scale of census districts or neighbourhoods rather than whole cities or departments) should help to improve exposure estimates. Unfortunately, lack of exposure data often limits this approach.

Box 4.5. Economic costs of different types of exposure assessment

Consider a study of the relationship between indoor NO_2 and the respiratory health of children. Respiratory health status is measured by questionnaire and lung function test. Associated costs per individual study subject are US$ 10 for the questionnaire (printing, mailing, data entry) and US$ 40 for the lung function test (technicians' salaries, travel cost, depreciation of equipment). To this must be added a further US$ 50 per subject for the overall cost of running the study (research associate salary, computing and reporting costs). C_i would therefore amount to US$ 100. (C_i =100)

For exposure assessment, the choice is between a questionnaire to collect data on indoor NO_2 sources, etc., and two home visits to install and collect passive NO_2 samplers. Associated costs per individual study subject are US$ 10 for the questionnaire (printing, mailing, data entry) and US$ 130 for the home visits. Additionally, preparation and analysis of the passive samplers costs US$ 10/sample, which amounts to US$ 30 per individual study subject if three samples are taken in each home.

C_a is therefore US$ 10, and C_t US$ 160 (C_t =130+30=160). A conservative estimate of the correlation between indoor NO_2 and home characteristics is 0.7. Entering the above values in the equation:

$$ARE_{a,t} = r^2 [(C_i + C_t) / (C_i + C_a)]$$

gives a value of: 0.49 [(100 + 160) / (100 + 10)] = 1.16, suggesting that, in this case, it would be more efficient to assess exposure by questionnaire than by taking indoor NO_2 measurements.

Use of existing data

Use of existing environmental exposure databases for epidemiological studies has been receiving increasing attention. Routinely collected environmental data are widely available in most countries. The data collected by some national and local authorities are used to create databases for the Global Environment Monitoring System (WHO, 1990, 1991a; UNEP/WHO, 1993). These provide data on pollution levels at specific sites, which can then be used to characterize average exposures for geographical regions. Sexton et al. (1992) provide an overview of 67 different exposure-related databases sponsored by the US Federal Government. However, they report that few contain data of actual measurements of exposure or dose in human populations. Matanoski et al. (1992) point out that existing exposure databases were created mostly for regulatory purposes, rather than primarily for use in public health or for epidemiological studies. Nevertheless, they believe that, with various targeted improvements, such databases could become useful tools for epidemiological studies.

The usefulness of data from existing databases for assessing exposure obviously depends on how well the existing data correlate with the true exposure of the subjects. For example, measurements of ambient particulate matter collected once every six days at a specified location could be used to assess long-term exposure of that community in a study comparing the health status of different communities. However, for this approach to be effective, the exposure within that community would have to be relatively homogeneous, and the air pollution levels of the communities significantly different. In addition, the correlation between the surrogate air

pollution measured at one location, and the true exposure of individuals residing in that location would have to be reasonably high. As an example, Dockery et al. (1993a) followed cohorts of adults living in six US communities, for approximately 16 years. The communities were characterized by different levels of long-term exposure to airborne particulate matter. Levels of PM_{10} met US Air Quality Standards in each of these communities, but mortality of subjects living in the most polluted communities was higher than among subjects living in the cleanest communities.

Prospective versus historical assessment

The environmental epidemiologist is often interested in the effect of environmental exposures on chronic diseases such as lung cancer. The time lag between exposure and other health effects for many pollutants may be of considerable interest too. Both the induction period and the latency period must be accounted for when determining the relevant exposure period. The induction period is the interval from exposure to initiation of the disease. The latency period is the time that lapsed between initiation of the disease in question and the onset of symptoms; in other words, health outcomes currently being observed may be due to exposures that occurred many years or even decades earlier. In such cases, exposure must be assessed retrospectively. This can be problematic, however, since historic data on pollution levels are often especially sparse.

In occupational epidemiology, "job-exposure matrices" have been developed to assist investigators in assigning specific exposures retrospectively to individual workers. These matrices are based on the basis of descriptions of the jobs that they have held. Methods for reconstructing environmental exposures retrospectively on the basis of residential histories are still in their infancy, though. Nevertheless, interesting examples of the latter can be found in recent studies on the development of lung cancer in relation to exposure to radon (Svensson et al., 1989) and air pollution (Jedrychowski et al., 1990).

Svensson et al. conducted a case–control study on lung cancer in women in Sweden, comparing the radon concentrations in their previous dwellings for cases and controls. The study included 210 incident female lung cancer cases, 209 population controls, and 191 hospital controls. The 610 subjects had lived for more than 2 years in a total of 3518 dwellings. Radon measurements were collected for a random sample of 303 dwellings and used to estimate lifelong radon exposure for all cases and controls. The results were analysed together with data on smoking. (See Table 4.16 for the main results.) The data show that lung cancer risk increased with increasing exposure to radon, although the increase was confined largely to the smokers. This study represents one of the few systematic attempts in the literature to make quantitative, retrospective estimates of exposure to an environmental agent.

Jedrychowski et al. (1990) compared 1579 lung cancer deaths in Cracow, Poland, as reported in the death register for the period of 1980–1985, to a reference population of 1491 other deaths (excluding deaths from respiratory causes). The deaths were frequency-matched for age and sex. Next-of-kin were interviewed to obtain information on individual smoking habits, occupation and last residence prior to death. Subjects' residences were classified in air pollution exposure categories based on existing records of total suspended particulate and sulphur dioxide for the relevant area. After adjustment for smoking and occupational exposures, the investigators found

that in Cracow air pollution may increase lung cancer risk, acting synergistically with smoking and occupational exposure. This is another example of a study in which the investigators tried to obtain historical data on environmental exposure and also on exposure to two different major confounders. The study was more limited than the Swedish radon study, however, since the investigators did not attempt to obtain information about previous places of residence.

Table 4.16. Relative risk estimates for smoking and radon exposure

	Estimated radon exposure level		
Smoking	< 4500 Bq/m^3	4501–6000 Bq/m^3	> 6000 Bq/m^3
Never	1.0 (reference)	1.4	0.9
1–10 cigarettes/day	2.3	4.8	6.5
> 10 cigarettes/day	6.8	12.3	15.9

Source: adapted from Svensson, et al., 1989.

Validity and precision

As discussed in Chapter 3 on pp. 57, the **validity** of an exposure measure refers to the agreement between this measure and the true exposure. Low validity is due mainly to systematic error in the exposure assessment. If the indoor presence of smokers determines individual exposure to dust in air, for example, the validity of the dust concentration measured at one central monitoring site in an urban area as a measure of true exposure to dust will be low. This is because it will not correlate strongly with personal exposure and will *under*estimate the dust exposure of subjects living with smokers. Moreover, it may *over*estimate the dust exposure of subjects who do not live with smokers, since it will not take into account the protection from dust in ambient air that they gain from being indoors. In addition, systematic measurement error may occur when using personal samplers; for example, if the sampling device is not calibrated correctly or if the resulting specimen is not handled carefully or if correct procedures are not followed during laboratory analysis.

Much attention has also been paid to the possibility of systematic measurement error by measuring equipment. However, random error due to a poor correlation between a measure of exposure and true exposure may be a more serious problem than a systematic difference between measured exposure and true exposure. In the latter case, the power of a study to detect a relationship between exposure and disease is not compromised. Theoretically, it would be possible to correct for the systematic error. If a personal sampling pump is calibrated incorrectly so that it samples 1.9 litres of air per minute instead of 2.0 litres per minute, the calculation of air concentration based on the sampling data will be incorrect. However, analysis of the exposure–response relationship would not necessarily be incorrect since each measurement would be incorrect to the same degree. Furthermore, if the calibration error were to be identified (and it should be if the pumps are checked regularly), it would be possible to adjust the sampling results mathematically to adjust for the error. On the other hand, if the sampling pump were

unreliable and the air sampling rates variable (i.e. imprecise) and not predictable, adjustments for error would not be possible.

The **precision** of a measure of exposure refers to its repeatability. Low precision is due to random error, and often related to the environmental sampling or the instrumentation and analytical procedures. However, sampling and analysis errors are often much less significant than the true variability of exposure measures in time and space. Assessment of the precision of a measure of exposure calls for prior definition of the spatial and temporal characteristics of the true exposure variable. Exposure is often measured in much more limited domains of time and space than those in which true exposure occurs. Repetition of exposure measurements in time and space is therefore needed to estimate variability of the true exposure.

Efforts to reduce misclassification of exposure should first concentrate on quality control of the exposure measurement process. (This topic is discussed in Chapter 8.) Another strategy is to obtain multiple measurements of exposure for all study subjects. When estimates of the within- and between-subjects components of variance can be extracted from published data or from a pilot study of the prospective study population, a calculation can be made to optimize the number of measurements per person (i.e. repeating measurements on the same person to obtain a consistent measure), relative to the total number of subjects to be studied. However, repeating exposure measurements may be costly and place undue burdens on the study subjects.

Estimates of potential confounders can also be misclassified. If there is non-differential misclassification of a confounder, the statistical adjustment for confounding may be incomplete. Misclassification of exposure variables and confounders should therefore be considered simultaneously and with equal emphasis.